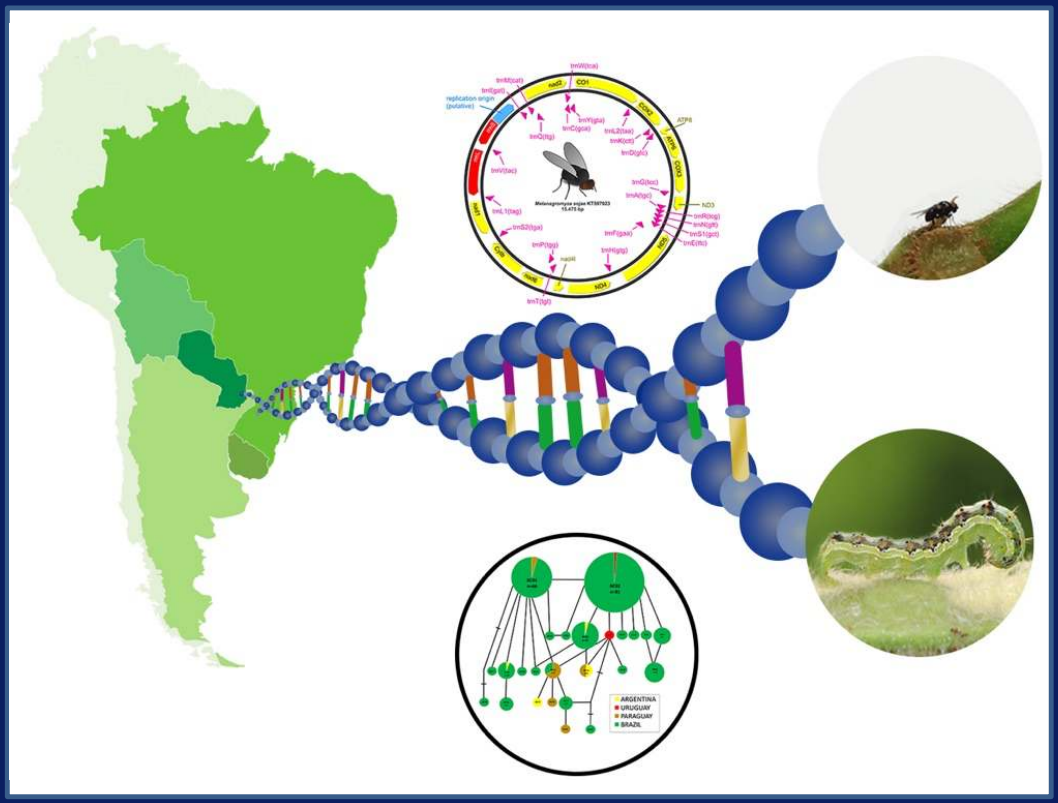


# Molecular tools to identify and study invasive pests in Brazil

Jonas André Arnemann

Universidade Federal de Santa Maria
1960

UNIVERSITEIT GENT

FACULTY OF BIOSCIENCE ENGINEERING

Thesis submitted in fulfillment of the requirements for the degree of
Doctor (PhD) in Applied Biological Sciences at the faculty of Bioscience Engineering of Ghent University, with co-tutelage with the Post-Graduate Program of the Federal University of Santa Maria

I dedicate this thesis to my parents Nelson and Helena, and to my brother Jardel.

*"Be kind; everyone you meet is fighting a hard battle."*

John Watson

*"Talvez não tenha conseguido fazer o melhor, mas lutei para que o melhor fosse feito. Não sou o que deveria ser, mas Graças a Deus, não sou o que era antes."*

Marthin Luther King

**Promotors:**

**Prof. dr. ir. Guy Smagghe**
Department of Crop Protection
Faculty of Bioscience Engineering
Ghent University, Ghent, Belgium

**Prof. Dr Jerson V. Carús Guedes**
Department of Entomology
Federal University of Santa Maria
Santa Maria, Rio Grande do Sul, Brazil

**Chair of the examination committee:**

**Prof. Dr. Élgion Lúcio da Silva Loreto**
Department of Biology
Federal University of Santa Maria
Santa Maria, Rio Grande do Sul, Brazil

**Members of the examination committee:**

**Prof. Dr. Rodrigo J. Seminoti Jacques**
Department of Soil Science
Federal University of Santa Maria
Santa Maria, Rio Grande do Sul, Brazil

**Dr. Wee Tek Tay, PhD**
Senior Research Scientist
CSIRO Health and Biosecurity Flagship
Canberra, ACT, Australia

**Prof. Dr. Juliano Tomazzoni Boldo**
Department of Biology
Federal University of Pampa-Unipampa
São Gabriel, Rio Grande do Sul, Brazil

**Prof. Dr. ir. Peter Bossier**
Department of Animal Production
Faculty of Bioscience Engineering
Ghent University, Ghent, Belgium

| | |
|---|---|
| **Dean:** | **Prof. dr. ir. Guido van Huylenbroeck** |
| **Rector:** | **Prof. dr. Anne De Paepe** |

**Dutch translation of the title:**

Het gebruik van moleculaire tools voor de identificatie van nieuwe invasieve plaagorganismen in Brazilië.

**Portuguese translation of the title:**

Uso de ferrmentas moleculares para identificar e estudar pragas invasivas no Brasil.

**Front cover:**

Top: xxx

**Please refer to this work as follows:**

Arnemann, JA. (2015). Molecular tools to identify and study invasive pests in Brazil. PhD Thesis. Ghent University, Ghent, Belgium.

ISBN: 978-90-5989-841-7

This study was financially supported with grants from the CAPES, UFSM and CSIRO.

# Table of Contents

# List of abbreviations

APHIS - Animal and Plant Health Inspection Service

CBD - Convention on Biological Diversity

COI - Cytochrome Oxidase subunit I

CONAB - Brazilian Companhia Nacional de Abastecimento

COSAVE - Comite de Sanidad Vegetal

Cyt *b* - Cytochrome *b*

DNA - Deoxyribonucleic Acid

DSV - Department of Plant Health

EPPO - European and Mediterranean Plant Protection Organization

FAO - Food and Agriculture Organization of the United Nations

FAOSTAT - The Statistics Division of FAO

gDNA - Genomic DNA

iBoL - International Barcode of Life

MAPA - Brazilian Ministry of Agriculture, Livestock and Food Supply

MtDNA - Mitochondrial DNA

NGS - Next-Generation Sequencing

PCGs - Protein Coding Genes

PCR - Polymerase Chain Reaction

RPPO - Regional Plant Protection Organization

SENAVE - Servicio Nacional de Calidad y Sanidad Vegetal y de Semillas

SNP - Single Nucleotide Polymorphism

SSF - Soybean Stem Fly

SSM - Soybean Stem Miner

USDA - United States Department of Agriculture

VIGIAGRO - Brazil's International Agricultural Defense System

# CHAPTER I

# MOLECULAR TOOLS TO IDENTIFY AND STUDY INVASIVE PESTS IN CONSE SUL OF AMERICA

## 1.1 Biosecurity, food security and agriculture

Biosecurity measures seek to protect the economy, environment and society from accidental or intentional introductions of invasive pests that can put the system at risk (Venette 2015). Sustainable development depends on the ability of a country to protect its ecosystems, economy and public health.

Over the past two centuries, the human population has grown seven-fold and demographers anticipate the addition of 2-3 billion more during the twenty-first century (Fedoroff 2015). The Food and Agriculture Organization of the United Nations (FAO) has estimated that amount of food produced needs to increase by 70% by 2050 (FAO 2009), not just because of the increase in population but because of changes in consumption patterns. In many countries the best land is already under cultivation and preserving what remains of our planet's rich biological heritage by leaving more land unploughed is a growing priority. Indeed, modelling food scenarios reveals that within just a few decades, the planet's natural resources will be insufficient to support developed-world consumption patterns (Odegard and Van der Voet 2014).

To increase the food production, the widespread introduction and distribution of economically desired plants and animals has helped to transform native flora and fauna around the world into what has been described as an indistinct mix of species (Mooney and Hobbs 2000). To obtain and to maintain high agricultural productivity will greatly depend on continued innovation necessary for controlling weeds, diseases, insects, as they evolve resistance to different control measures, or as new pest species emerge or became dispersed to new regions (Godfray et al. 2010). The vast majority of such species are necessary to maintain high agricultural productivity, but a small

percentage of them can spread rapidly beyond their introduced areas and become invasive species (Ziska et al. 2011).

As used by Guillemaud et al. (2011) 'an invasion may be considered to have occurred when a group of individuals has been introduced into a new area, in which they have established themselves, increased in number and spread geographically'. Prevention of biological invasions, as opposed to remedial eradication of invasive species, represents the most cost-effective and perhaps the only hope for stemming the current homogenization of the world's biota (Mack et al. 2000; Miller et al. 2005).

Biological invasions constitute a major environmental change driver, affecting conservation (e.g., Gurevitchemail and Padilla 2004), human health (e.g., invasive mosquitoes, Juliano and Philip 2005) and agriculture (e.g., in Brazil introduction of Asian citrus psyllids *Diaphorina citri* (Hemiptera: Liviidae), Guidolin et al. 2014 and the cotton boll weevil, *Anthonomus grandis* in 1983, Barbosa et al. 1983).

The Brazilian success in agribusiness in recent decades is due to its extensive arable lands, favourable climate, intensive use of basic inputs and mechanization, expanded farm credit, price support and export incentive policies and in particular, research by government agencies and universities throughout the country (Barros et al. 2009; Oliveira et al. 2013). As invasive insect pests bring a significant and unprecedented threat to the productivity of agriculture, measures to control these pests must be taken, including actions to prevent the spread of exotic species to new areas and the development of new management systems (Liebhold and Tobin 2008).

## 1.2 Invasive pests in South America and Brazil

The resources required to prevent an invasive species establishing in an area are usually significantly lower than those needed for eradication, containment, long-term control, or the consequences of doing nothing (Maynard and Nowell 2009). Pest occurrence in neighbouring countries and international trade partners should be used to determine the need of reviewing established policies (Lopes-da-Silva 2014). In Latin America the 'Comite de Sanidad Vegetal' (COSAVE), which is a Regional Plant Protection Organization (RPPO), set up under the International Plant Protection Convention (IPPC). It operates as an intergovernmental organisation to coordinate and facilitate cooperation to solve phytosanitary problems of common interest to its member countries and strengthen regional integration. Argentina, Brazil, Chile, Paraguay and Uruguay are members. Bolivia is also participating in the activities of the organization and on the later stages of the formal accession process (COSAVE 2015).

The 'COSAVE' is directly involved with main issues around invasive pests, such as risk analysis, plant quarantine and phytosanitary certification. According to the definitions adopted by the International Convention on Biological Diversity (CBD 1992), a species is considered exotic (or introduced) when located at a site outside their natural range because of an introduction mediated by human actions. If the introduced species can reproduce with fertile offspring having a high probability of surviving in the new habitat, it is considered established. If the established species expand their distribution in the new habitat, threatening the native biodiversity, it is now considered an exotic invasive species (Leão et al. 2011).

These definitions provided by the CBD are used as a reference for the legal foundation and becomes part of public policy of the signatory countries of the

Convention, including Brazil. It means that the country must "prevent the introduction and control or eradicate alien species which threaten ecosystems, habitats or species" (Art. 8 in the CBD, 1992). This article was transposed into the Law of Environmental Crimes (art. 61 of the Federal Law No. 9.605 / 98), which considers environmental crime the spread of diseases or pests or species that may cause damage to agriculture, livestock, fauna, flora, or to ecosystems (CBD 1992; Leão et al. 2011).

Brazil may have an efficient system of risk management and prevention to invasive pests (e.g. VIGIAGRO/MAPA) but if a neighbouring country (and Brazil has 10 neighbouring countries: Uruguay Argentina, Paraguay, Bolivia, Peru, Colombia, Venezuela, Guiana, Suriname and French Guiana) detects a new invasive pest, the best policy is to be prepared. This includes implementing border monitoring for early detection and a detailed mitigation plan. The Brazilian government had recently added organic material detectors at its main ports and airports, technology that Argentina and Chile already had. Further measures were conducted in conjunction with the national intelligence agency when the World Cup started in June 2014 (Reuters 2014). According to Lopes-da-Silva (2014) there are at least 150 quarantine pests regulated by Brazil that are already in other South American countries.

Brazilian growers have faced many problems with invasive pests, e.g., *Tuta absoluta* (Muszinski et al. 1982); *Thrips palmi* (Monteiro et al. 1995); the cotton boll weevil *Anthonomus grandis* (Barbosa et al. 1983); the whitefly *Bemisia tabaci* MEAM 1 (previously known as ‘B biotype’) in 1991 (Lourenção and Nagai 1994); the Asian Soybean Rust *Phakosphora pachyrhizi* in 2001 (Yorinori et al. 2005); and the Old World cotton bollworm *Helicoverpa armigera* in 2013 (Czepak et al. 2013; Tay et al. 2013).

Recently, new invasive pest had also been confirmed in Brazil, as the Soybean Stem Fly (SSF) *Melanagromyzae sojae* (Arnemann et al. 2015; Chapter II and III).

The introduction and establishment of invasive pests can cause direct economic losses through negatively impacting on access to agricultural product export markets (e.g., loss of 'pest-free' status; Follett and Neven 2006) and increase damage to crops, and resulting in the need to develop and adopt control measures. Genetic tools are effective for rapidly confirming the presence/absence of certain invasive/exotic pest species so that potential economic losses (e.g., embargos on phytosanitary products) may be avoided. Furthermore, they can assist with effective monitoring and management of exotic pest species once identified and appropriate markers developed.

Managers seeking to control and/or to mitigate (e.g., through better monitoring/inspection programs) the spread of invasive pest species will benefit from information that helps to identify source populations and incursion routes. Effective genetic tools can also provide information on potential origin(s) of an invasive species, determining if introduction was intentional or as a result of unintentional released and/or through escape from captivity (Hunter 2012). This may also have implications for identifying the route of entry and assist in preventing further invasions from the same source.

Currently, Brazil's International Agricultural Defense System (VIGIAGRO) which is the responsibility of the Department of Plant Health (DSV), part of the Brazilian Ministry of Agriculture, Livestock and Food Supply (MAPA-Brazil) with the DSV being the national plant protection organization (NPPO) in Brazil, as established by Article IV of the International Plant Protection Convention (Oliveira et al. 2013).

Brazil has a high risk of exotic species arriving and establishing, as Brazilian agriculture is characterized by large export and import flows, allied with the country's vast landmass and 15.7 thousand kilometres of borders with 10 neighbouring countries in South America provide numerous entry points (Oliveira et al. 2012; IBGE 2015). Many exotic species around the world are insects (such as the soybean stem fly, Chapter II and III) as their small size and great ability to survive under unfavourable conditions during the transport and storage of agricultural products and plant propagation materials make them more likely to arrive alive. In addition to their ability to arrive in a country they also often posses a high capacity for dispersal and colonization (e.g. high reproductive output and multivoltine) in new environments after arrival (Kiritani and Yamamura 2003; Bounfour et al. 2005).

The Department of Plant Protection from the Department of the Ministry of Agriculture, Livestock and Supply (MAPA) is responsible for verifying the quarantine importance of pests and establish regulations for the transit of vegetable hosts of quarantine pests. The detection of a quarantine pest may occur during import or in the field. When found in the field, MAPA guidelines requires the detailed description of how the pest was detected, on which host, the damage observed, the method and the technique of identification used, besides locations of detection. MAPA also uses information generated by researchers to support the pest presence in Brazil (e.g., Chapter II and Chapter III) (Filho and Zucchi 2015).

### 1.3 The Soybean Stem Fly (SSF), *Melanagromyza sojae* (Zehntner)

In tropical and subtropical areas of Asia and Africa, species of Agromyzidae flies as *Melanagromyza sojae*, *Melanagromyza dolichostigma*, *Ophiomyia phaseoli*,

*Ophiomyia spencerella* and *Ophiomyia centrosematis* are often found in soybean (*Glyicne max)* and bean crops and are considered important pests. The soybean stem fly (SSF) *M. sojae* (Zehntner) belong to the family Agromyzidae and is highly polyphagous (i.e., attacking plants such as *G. max, Phaseolus vulgaris, Pisum sativum, Vigna angularis,* and other members from the Fabacea Family; Dempewolf 2004) and is one of the major issues for cultivation of soybean in Northern India (e.g., Wang and Gai 2001), in Iran (e.g., Ziaee 2012), Russia (Strakhova et al. 2013), in Asia (e.g., China, Wang and Gai 2001); Nepal (e.g., Thapa 2012), and in parts of South East Asia (e.g., Indonesia, Van Den Berg et al. 2008), and is a potential pest for Brazil (Hirose and Moscardi 2012) and for northern Australia (Shepard et al. 1983). SSF may also represent a significant pest in North America although its presence has not been reported.

Wang (1979) shows that the egg stage of SSF in soybean takes between 2 and 7 days, the 3 larval instars around 7.7 days, and the pupal stage between 6 and 12 days. The adult life-span is 19 days and females lay an average of 171 eggs. In Iran, Ziaee (2012) described SSF having four or five generations per year and pupae overwintering in the stem.

There are many reports regarding damage to soybean fields by SSF, from those that reported 100% infestation of plants and significant losses, to losses of 2% in grain yield. This variation may depend on the region of the occurrence of the pest, the plants nutritional status, the soybean cultivar used, the sowing date and the use of control measures (Savajji 2006).

An infestation by SSF in soybean field was recorded in India in 2010, with up to 73% of plants infested with 48% stem tunnelling, and the authors mentioned that an

increase in rainfall, the minimum temperature and relative humidity caused declines in the population (Yadav et al. 2015). Other reports from India indicated that SSF infested 100% of plants and tunnelled up to 70% of stem length (Singh and Singh 1990; Singh and Singh 1992). Talekar (1989) pointed to a reduction in soybean plant height, leaf area, dry matter of soybean and nodulation by *Rhizobium*; in additional to decreased leaf numbers, and grains and soybean yield when *G. max* were under attack of SSF.

There is no genetic information about *M. sojae* in GenBank or iBoL (consulted 25th Aug 2015), which would enable genetic comparison, species confirmation and assist with genetic diversity studies. *Bemisia tabaci*, e.g., has approximately 254.209 sequences in GenBank (genomic DNA/RNA=7.470 and mRNA=139.685, 25th Aug 2015), which permits many molecular studies (e.g., in 13 September 2015, up to 106 papers in PubMed using '*Bemisia tabaci* genetic diversity' as key words) and also molecular confirmation. The lack of SSF information in GenBank and iBoL prevents molecular identification and makes genetic population studies much more difficult.

### 1.4 *Helicoverpa armigera* occurrence in Brazil

The confirmation of the presence and invasion of *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae) in south and central America (in Brazil in 2013 by Czepak et al. 2013 and Tay et al. 2013; in Paraguay in 2013 by SENAVE 2013; in Argentina in 2014 by Murua et al. 2014) has potential serious implications in terms of the management of insect pests in the main agricultural crops cultivated in these areas. In last July (2015) USDA/APHIS and Florida Department of Agriculture and Consumer Services (FDACS), also confirmed the detection of *H. armigera* in the USA territory.

The literature on *H. armigera* biology is extensive (e.g., Coaker 1959; Hardwick 1965; Raulston et al. 1986; Fitt 1989). Worldwide, *H. armigera* cannot survive winter north of 40° latitude and seasonal populations are sustained by immigration (Hardwick 1965).

*H. armigera* eggs are laid singly on various plant structures such as leaves, buds, flowers, fruit and stems (Zalucki et al. 1986). The egg incubation period takes an average of 3.3 days (Ali and Choudhury 2009), and the duration of larval development in *H. armigera* is very variable, and can be extended to six or seven instars instead of the more normal five or six instars (Zalucki et al. 1986). Where hosts are available in the tropics (e.g., some regions in Brazil) *H. armigera* may breed continuously, completing a generation in 28-30 days and passing through 10-11 generations per year (Coaker 1959; Raulston et al. 1986; Fitt 1989).

Fitt (1989) listed the four major factors leading to the success of *H. armigera* as a worldwide pest: (i) high polyphagy (i.e., attacking a diverse array of plant species in many plant families worldwide), affording a great potential for population persistence and increase; (ii) the ability to undertake local and interregional movements (e.g., the vertical take-off fligh, which carries them above the flight boundary layer and enables them to undertake migratory movement in upper wind systems, at altitudes of up to 1-2 km); (iii) the ability to enter a facultative diapause as pupae, i.e., becoming more tolerant of extreme climatic conditions; (iv) the high fecundity, which combined with a short generation time, gives this species a high capacity for population increase.

The wide spread presence of *H. armigera* in Brazil led to the believe that this pest was present in Brazil before 2013 but not reported, due in part, to its morphological similarity to other related *Heliothinae* species (Pomari-Fernandes et al.

2015) such as the closely related *H. zea* (e.g., Behere et al. 2007; Cho et al. 2008). Through sequencing a portion of the mitochondrial genes Cytochrome Oxidase subunit I (mtDNA COI) and Cytochrome *b* (mtDNA cyt *b*) of suspected field-collected individuals in Brazil enabled the presence of this species in the New World to be confirmed at the molecular level, while the inclusion of additional nuclear markers enable high genetic diversity in the invasive population to be concluded (Tay et al. 2013). A well-established and characterized DNA database such as the mtDNA COI gene (e.g., Behere et al. 2007; Behere et al. 2008) underpinned the efficiency for molecular confirmation of invasive exotic species at national and transnational border scales (Tay et al. 2013; Mastrangelo et al. 2014; USDA/APHIS 2015).

### 1.5 Mitochondrial DNA markers for pest studies

Using mitochondrial sequence data to understand invasion patterns has advantages over use of nuclear markers, such as microsatellites, especially because the mitochondrial genome is typically present in higher copy number in cells (Rollins et al. 2011). Mitochondrial DNA variation should reflect more recent demographic events than nuclear data because the lower effective population size of mtDNA results in a decrease in fixation time for new mutations (Rollins et al. 2011).

Mitochondrial genes are widely used to study population differentiation at and below the species level (Avise 2000), have a smaller effective population size than do nuclear genes and, when comparing neutral loci, are expected to track differentiating populations more rapidly (Moore 1995; Palumbi et al. 2001). Traditionally the mtDNA has been considered as almost exclusively inherited through the maternal lineage (e.g., Giles et al. 1980; Crozier 1990). More recently, this mode of inheritance has

increasingly been challenged (e.g., Nunes et al. 2013; Dokianakis and Ladoukakis 2014) due to greater detection rates of paternal mtDNA leakage. Other issues such as detection of mitochondrial genome recombination, and heteroplasmy (i.e., presence of different mtDNA molecules within an individual, e.g., see Robinson et al. 2015) events in diverse organisms (e.g., see reviews by Rubinoff et al. 2006; Krishnamurthy and Francis 2012) have also been reported.

Incorporating ecological data, behavioural traits, and morphological characters with DNA characterisation (i.e., the mtDNA COI gene) will enable meaningful species delineation including identification of cryptic species which may be of significant implications to insectice resistance and invasive species management strategies (e.g., Hebert et al. 2004; Scheffer et al. 2006; Rubinoff et al. 2006; Chapters II and III). Developing effective PCR markers to take advantage of existing DNA database, such as from the international consortium effort to 'barcode' biodiversity (e.g, iBoL) or GenBank, can help to advance research into understanding diverse research questions including assisting with identifying invasive pest species. It is necessary to keep in mind that the integrity of DNA sequence identity especially those generated from public input (e.g., the iBoL database) as inaccuracy are occasionally introduced and can affect interpretations of phylogenetic relations between species (e.g., Hebert et al. 2004; Smith et al. 2005).

The use of insect mtDNA as a molecular marker for studies of population genetic is well-reported/studied on account of ease in manipulation, rapid mutation rate, general lack of significant recombination, and availability of universal primers (e.g., for the sugarcane borer *Diatraea saccharalis,* Silva-Brandão et al. 2015; for *Heliothis virescens*, Albernaz et al. 2012). Some studies (e.g., for the red tomato spider

mite, *Tetranychus evansi,* Boubou et al. 2011) demonstrated the extent to which molecular markers can aid in the study of an emerging pest in the earliest stages of the invasion process and thus help to address questions related to management and biosecurity. Some literature mentioned that regulations should also take into account the fact that some genotypes of an invasive species may differ in invasive potential (Allendorf and Lundquist 2003).

Perhaps the greatest challenge is to encourage the integration of molecular information, combining the technical and scientific advances of molecular tools to analyse DNA with taxonomy, to better understand the invasion patterns and gene flow of new invasive pests.

### 1.6 Using molecular tools to help confirm invasive pests and study population dynamics

By integrating traditional system of species identification (e.g., via morphology) with new technological advances such as individual gene-based sequence (e.g., Sanger sequencing method) as well as NGS (e.g., whole genome sequencing) has greatly propelled the speed in which novel pest incursions are being confirmed (e.g., Tay et al. 2013; Chapters II and III; see also Kirk et al. 2013). Although with wide adoption of NGS approaches due to significant cost reduction for NGS methods in recent times (Carlson 2011), it is perhaps important to keep in mind that without well-founded basic taxonomic research, the rapid discovery of novel genome contents would be meaningless (Rubinoff et al. 2006).

Recent confirmation of the high profile incursions by the Old Wold cotton bollworm *Helicoverpa armigera* in South and North American continents (e.g., Tay et

al. 2013) demonstrated the benefit and need to accurately and rapidly identify invasive pests. Although molecular methods via PCR, RFLP and/or sequencing to accurately differentiate *H. armigera* from *H. zea* have been developed and reported by Behere et al. (2008), the adoption of these methods was low in Brazil (e.g., Queiroz et al. 2013). Furthermore, delay in identifying suspected *H. armigera* in Brazil during the initial outbreaks also likely significantly reduced opportunities to effectively manage the incursion, and may have contributed to the rapid population growth and new outbreaks across South American's agricultural landscapes. Although the Brazilian *H. armigera* is now likely a stable, self-sustaining population, its contributions as a source population for new incursions across the New World remained to be accurately assessed (e.g., Chapter III), despite the spatial-climatic modelling of population expansion across the North and South America (Kriticos et al. 2015).

In this thesis the Agromyzidae fly *Melanagromyza sojae*, which is a potentially serious insect pest species of soybean *Glycine max* (L.) in Brazil, was first identified via larval morphological characters that enable rapid molecular characterisation of its mitochondrial genome via the NGS method (Arnemann et al. 2015; Chapter II), and enable establishing of a mtDNA COI partial gene sequence database that also included matching of mtDNA haplotypes with previously confirmed *M. sojae* specimens reported in Australia. This has allowed the world's first study into the SSF population genetic diversity from Southern Brazil to be carried out and has established a baseline for future work on this important pest species (Chapter III). The value of a well-surveyed global *H. armigera* mtDNA COI haplotypes database to understanding patterns of incursions is demonstrated in Chapter IV, where spatial distribution

patterns of *H. armigera* genetic diversity lends weight to support the hypothesis of multiple independent incursions of *H. armigera* into South America in recent times.

## 1.7 References

Albernaz KC, Silva-Brandao KL, Fresia P, Cônsoli FL, Omoto C. 2012. Genetic variability and demographic history of *Heliothis virescens* (Lepidoptera: Noctuidae) populations from Brazil inferred by mtDNA sequences. *Bulletin of entomological research* 102: 333-343.

Ali A, Choudhury RA, Ahmad Z, Rahman F, Khan FR, Ahmad SK. 2009. Some biological characteristics of *Helicoverpa armigera* on chickpea. *Tunisian Journal of Plant Protection* 4: 99-106.

Allendorf FW, Lundquist LL. 2003. Introduction: population biology, evolution, and control of invasive species. *Conservation Biology* 17: 24-30.

Arnemann JA, Walsh T, Gordon K, Brier H, Guedes JVC, Tay WT. 2015. Complete mitochondrial genome of the soybean stem fly *Melanagromyza sojae* (Diptera: Agromyzidae). *Mitochondrial DNA* (Accepted 09-Sept-2015, In press).

Avise JC. 2000. *Phylogeography, the History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts.

Barbosa S, Braga SR, Lukefahr MJ, Beingolea GO. 1983. Relatorio sobre a ocorrencia do bicudo do algodoeiro, *Anthonomus grandis* Boheman “Boll Weevil”, no Brasil e recomendacoees para a sua erradicacao. Documentos 21. EMBRAPA CNPA, Campina Grande, Brazil.

Barros GSC, Spolador HFS, Bacchi MRP. 2009. Supply and demand shocks and the growth of the Brazilian agriculture. *Revista Brasileira de Economia* 63: 35-50.

Behere GT, Tay WT, Russell DA, Heckel DG, Appleton BR, Kranthi KR, Batterham P. 2007. Mitochondrial DNA analysis of field populations of *Helicoverpa armigera* (Lepidoptera: Noctuidae) and of its relationship to *H. zea*. *BMC Evolutionary Biology* 7: 117-127.

Boubou A, Migeon A, Roderick GK, Navajas M. 2011. Recent emergence and worldwide spread of the red tomato spider mite, *Tetranychus evansi*: genetic variation and multiple cryptic invasions. *Biological Invasions* 13: 81-92.

Bounfour M, Jebbour F, Wadjinny J. 2005. Biological traits of invasive insect species harmful to Moroccan agriculture. In *Plant protection and plant health in Europe: introduction and spread of invasive species.* BCPC Conference, Humbolt University (DE), 2005-06-09/11: 95-100.

Carlson RH. 2011. Biology is technology: the promise, peril, and new business of engineering life. Cambridge, Mass. Harvard University Press.

CBD. 1992. Convention on Biological Diversity: Text and Annexes Convention on Biological Diversity Interim Secretariat, Geneva Executive Center, C.R 356, CH-1219 chatelaine, Switzerland.

Cho S, Mitchell A, Mitter C, Regier J, Matthews M, Robertson R. 2008. Molecular phylogenetics of heliothine moths (Lepidoptera: Noctuidae: Heliothinae), with comments on the evolution of host range and pest status. *Systematic Entomology* 33:581-594.

Coaker TH. 1959. Investigations on *Heliothis armigera* in Uganda. *Bulletin of Entomological Research* 50:487-506.

COSAVE. 2015. Comité de Sanidad Vegetal. Avaiable in: <http://www.cosave.org/pagina/bienvenidos-al-comite-de-sanidad-vegetal-cosave> Access in 24 Aug 2015.

Crozier RH. 1990. From population genetics to phylogeny: uses and limits of mitochondrial DNA. *Australian Systematics Botany.* 3: 111-24.

Czepak C, Albernaz KC, Vivan LM, Guimaraes HO, Carvalhais T. 2013. Primeiro registro de ocorrência de *Helicoverpa armigera* (Hübner) Lepidoptera: Noctuidae) no Brasil. *Pesquisa Agropecuária Tropical* 43: 110-113.

Dempewolf M. 2004. Arthropods of economic importance: Agromyzidae of the World. ETI- Information Services; Wokingham; UK (unpaginated).

Dokianakis E, Ladoukakis ED. 2014. Different degree of paternal mtDNA leakage between male and female progeny in interspecific *Drosophila* crosses. *Ecology and Evolution* 4: 2633-2641.

FAO. 2009. [cited 05 September 2015]. Available from: https://www.fao.org.br. Fedoroff NV. 2015. Food in a future of 10 billion. *Agriculture & Food Security* 4: 11.

Filho EV, Zucchi RA. 2015. Pragas Introduzidas no Brasil: insetos e ácaros. Piracicaba, FEALQ, 908 p.

Fitt GP. 1989. The ecology of *Heliothis* species in relation to agroecosystems. *Annual Review of Entomology* 34: 17-52.

Follett PA, Neven LG. 2006. Current trends in quarantine entomology. *Annual Review of Entomology* 51: 359-385.

Giles RE, Blanc H, Cann HM, Wallace DC. 1980. Maternal inheritance of human mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* 77: 6715-6719.

Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Robinson S, Thomas SM, Toulmin C. 2010. Food security: the challenge of feeding 9 billion people. *Science* 327: 812-818.

Guidolin AS, Fresia P, Cônsoli FL. 2014. The Genetic Structure of an invasive pest, the Asian Citrus Psyllid *Diaphorina citri* (Hemiptera: Liviidae). *PLoS ONE* 9(12): e115749. doi:10.1371/journal.pone.0115749.

Guillemaud T, Ciosi M, Lombaert E, Estoup A. 2011. Biological invasions in agricultural settings: Insights from evolutionary biology and population genetics. *Comptes Rendus Biologies* 334: 237-246.

Gurevitchemail J, Padilla DK. 2004. Are invasive species a major cause of extinctions? *Trends in Ecology & Evolution* 19(9): 470-474.

Hardwick DF. 1965. The corn earworm complex. *Memoirs of the Entomological Society of Canada* 97: 5-247.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes flugerator*. *Proceedings of the National Academy of Sciences of the United States of America* 101: 14812-14817.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. 2004. Identification of birds through DNA barcodes. *PLoS Biology* 2: 01-06.

Hirose E, Moscardi F. 2012. Insetos de outras regiões do mundo. IN: EMBRAPA-Soja: manejo integrado de insetos e outros artrópodes-praga. Brasília, DF: Embrapa, 859 p.

Hunter ME. 2012. Using Genetic Research to Inform Imperiled and Invasive Species Management. U.S. Geological Survey Fact Sheet 3110. [cited 05 September 2015]. Available from: <http://pubs.usgs.gov/fs/2012/3017/pdf/FS_2012-3017_Hunter-WEB.pdf>.

IBGE 2015. Instituto Brasileiro de Geografia e Estatitica. Noticias. [cited 01 September 2015]. Available from: http://www.ibge.gov.br/home/geociencias/geografia/default.shtm.

Juliano SA and Philip LL. 2005. Ecology of invasive mosquitoes: effects on resident species and on human health. *Ecology Letters* 8: 558-574.

Kiritani K, Yamamura K. 2003. Exotic insects and their pathways for invasion. In: Ruiz GM, Carlton JT, editors. Invasive species Vectors and management strategies. Washington: Island Press.

Kirk H, Dorn S, Mazzi D. 2013. Molecular genetics and genomics generate new insights into invertebrate pest invasions. *Evolutionary Applications* 6: 842-856

Krishnamurthy PK, Francis RA. 2012. A critical review on the utility of DNA barcoding in biodiversity conservation. *Biodiversity and Conservation* 21: 1901-1919.

Kriticos DJ, Ota N, Hutchinson WD, Beddow J, Walsh T, Tay WT, Borchert DM, Paula-Moreas SV, Czepak C, Zalucki MP. 2015. The potential distribution of invading *Helicoverpa armigera* in North America: Is it just a matter of time? *PLoS ONE* 10(7): e0133224.

Leão TCC, Almeida WR, Dechoum MS, Leão T, Ziller SR, Almeida WR. 2011. Espécies Exóticas Invasoras no Nordeste do Brasil: Contextualização, Manejo e Políticas Públicas. CEPAN-Instituto Hórus. Recife, 99p.

Liebhold AM, Tobin PC. 2008. Population ecology of insect invasions and their management. *Annual Review of Entomology* 53: 387-408.

Lopes-da-Silva M, Sanches MM, Stancioli AR, Alves G, Sugayama R. 2014. The role of natural and human-mediated pathways for invasive agricultural pests: A historical analysis of cases from Brazil. *Agricultural Sciences* 5: 634-646.

Lourenção AL, Nagai H. 1994. Surtos populacionais de *Bemisia tabaci* no Estado de São Paulo. *Bragantia* 53: 53-59.

Mack RN, Simberloff D, Lonsdale WM, Evans H, Clout M, Bazzaz FA. 2000. Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications* 10: 689-710.

Mastrangelo T, Paulo DF, Bergamo LW, Morais EGF, Silva M, Bezerra-Silva G, Azeredo-Espin AML. 2014. Detection and genetic diversity of a heliothine invader (Lepidoptera: Noctuidae) from north and northeast of Brazil. *Journal of Economic Entomology* 107: 970-980.

Maynard G, Nowell D. 2009. Biosecurity and quarantine for preventing invasive species. Acess in:< http://www.eubarnet.eu/category/libraries/biosecurity-environmental-aspects-and-gmos/page/2/>.

Miller N, Estoup A, Toepfer S, Bourguet D, Lapchin L, Derridj S, Kim KS, Reynaud P, Furlan L, Guillemaud T. 2005. Multiple transatlantic introductions of the Western Corn Rootworm. *Science* 310 992.

Monteiro RC, Zucchi RA, Mound LA. 1995. Record of *Thrips palmi* (Thysanoptera, Thripidae) in the State of Sao Paulo, Brazil. *Revista Agrícola* 70: 53-55.

Mooney HA, Hobbs RJ. 2000. Invasive species in a changing world. Island, Washington DC, p 457.

Moore WS. 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution* 49: 718-726.

Murúa MG, Scalora FS, Navarro FR, Cazado LE, Casmuz A, Villagrán ME, Lobos E, Gastaminza G. 2014. First record of *Helicoverpa armigera* (Lepidoptera: Noctuidae) in Argentina. *Florida Entomologist* 7: 854-856.

Muszinski T, Lavendowski IM, Maschio LMA, 1982. Constatacao de *Scrobipalpula absoluta* (Meyrick, 1917) (=*G norimoschema absoluta*) (Lepidoptera: Gelechiidae), como praga do tomateiro (*Lycopersicon esculentum* Mill.) no litoral do Parana. *Anais da Sociedade Entomológica do Brasi* 11: 291-292.

Nunes MD, Dolezal M, Schlotterer C. 2013. Extensive paternal mtDNA leakage in natural populations of *Drosophila melanogaster*. *Molecular Ecology* 22: 2106-2117.

Odegard IYR, van der Voet E. 2014. The future of food-scenarios and the effect on natural resource use in agriculture in 2050. *Ecological Economics* 97: 51-59.

Oliveira CM, Auad AM, Mendes SM, Frizzas MR. 2013. Economic impact of exotic insect pests in Brazilian agriculture. *Journal of Applied Entomology* 137: 1-15.

Palumbi SR, Cipriano F, Hare MP. 2001. Predicting nuclear gene coalescence from mitochondrial data: the three-times rule. *Evolution* 55: 859-868.

Pomari-Fernandes A, de Freitas Bueno A, Sosa-Gómez DR. 2015. *Helicoverpa armigera*: current status and future perspectives in Brazil. *Current Agricultural Science and Technology* 21: 1-7.

Queiroz PRM, Queiroz ÉSM,; Soares CM, Scoz LB, Thomazoni D, Soria, MF, Monnerat RG. 2013. Identificação molecular de *Helicoverpa armigera*: tecnologia a serviço dos cotonicultores de Mato Grosso. *Circular Tecnica IMA-MT*, n.6, 4p.

Raulston JR, Pair SD, Martinez FA, Westbrook JK, Sparks AN et al. 1986. Ecological studies indicating the migration of *Heliothis zea, Spodoptera frugiperda* and *Heliothis virescens* from northeastern Mexico and Texas, p204-220. In: Gatehouse AG. 1986. Migration in the African armyworm *Spodoptera exempta*: genetic determination of migratory capacity and a new synthesis. In: Insect Flight, Dispersal and Migration. W. Danthanarayana, pp. 128-144, Heidelberg, West Germany, Springer-Verlag.

Reuters 2014. New threat to Brazil's breadbasket: a pesky caterpillar. [cited 01 November 2015]. Available from: <http://www.reuters.com/article/2014/02/27/us-brazil-agriculture-caterpillar-idUSBREA1Q19L20140227>.

Robison GA, Balvin O, Schal C, Vargo EL, Booth W. 2015. Extensive mitochondrial heteroplasmy in natural populations of a resurging human pest, the Bed Bug (Hemiptera: Cimicidae). *Journal of Medical Entomology* DOI: <http://dx.doi.org/10.1093/jme/tjv055> First published online: 25 May 2015.

Rollins LA, Woolnough AP, Sinclair R, Mooney NJ, Sherwin WB. 2011. Mitochondrial DNA offers unique insights into invasion history of the common starling. *Molecular Ecology* 20: 2307-2317.

Rubinoff D. 2006. Utility of mithochondrial DNA barcodes in species conservation. *Conservation Biology* 4: 1026-1033.

Savajji, K. 2006. Biology and management of soybean stem fly *Melanagromyza sojae* (Zehntner) (Diptera: Agromyzidae). University of Agricultural Sciences. Master of Science thesis. 58pp.

Scheffer SJ, Lewis ML, Joshi RC. 2006. DNA Barcoding Applied to Invasive Leafminers (Diptera: Agromyzidae) in the Philippines. *Annals of the Entomological Society of America* 99: 204-210.

Senave. 2013. Senave en alerta tras ingreso de peligrosa plaga agrícola. [cited 05 September 2015]. Available from: <http://www.abc.com.py/edicion-impresa/economia/senave-en-alerta-tras-ingreso-de-peligrosa-plaga-agricola-629240.html>.

Shepard BM, Barrion AT. 1998. Parasitoids of insects associated with soybean and vegetable crops in Indonesia. *Journal of Agricultural Entomology* 15: 239-272.

Silva-Brandão KL, Santos TV, Cônsoli FL, Omoto C. 2015. Genetic diversity and structure of Brazilian populations of *Diatraea saccharalis* (Lepidoptera: Crambidae): implications for pest management. *Journal of Economic Entomology* DOI: http://dx.doi.org/10.1093/jee/tou040 First published online: 25 January 2015.

Singh KJ and Singh OP. 1992. Influence of stem tunneling by the maggots of *Melanagromyza sojae* (Zehn.) on yield of soybean. *Journal of Insect Science* 5: 198-200.

Singh OP, Singh KJ. 1990. Seasonal incidence and damage of *Melanagromyza sojae* (Zehnt.) on soybean. *Indian Journal of Plant Protection* 18: 271-275.

Smith MA, Fisher BL, Hebert PDN. 2005. DNA barcoding for effective biodiversity assessment of a hyper diverse arthropod group: the ants of Madagascar. *Philosophical Transactions of the Royal Society B* 360: 1825-1834.

Strakhova IS, Yefremova ZA, Tschirnhaus M, Yegorenkova EN. 2013. The parasitoid complex (Hymenoptera, Eulophidae) of leafminer flies (Diptera, Agromyzidae) in the middle Volga Basin. *Entomological Review* 93(7): 865-873. DOI: 10.1134/S0013873813070087.

Talekar NS. 1989. Characteristics of *Melanagromyza sojae* (Diptera: Agromyzidae) damage in soybean. *Journal of Economic Entomolgy* 82: 584-588.

Tay WT, Soria MF, Walsh T, Thomazoni D, Silvie P, Behere GT, Anderson C, Downes S. 2013. A brave new world for an Old World pest: *Helicoverpa armigera* (Lepidoptera: Noctuidae) in Brazil. *PLoS ONE* 8(11): e80134.

Thapa RB. 2012. Redescription of *Melanagromyza sojae* (Zehntner) from India and Nepal. *Nepalese Journal of Biosciences* 2: 64-70.

USDA/APHIS (Animal and Plant Health Inspection Service). Detection of Old World Bollworm (*Helicoverpa armigera*) in Florida - for information and action da-2015-43 July 30, 2015. [cited 05 September 2015]. Available from: <http://www.aphis.usda.gov/plant_health/plant_pest_info/owb/downloads/DA-2015-43.pdf>.

Van Den Berg H, Ankasah D, Hassan K, Muhammad A, Widayanto Ha, Wirasto Hb, Yully I. 1995. Soybean stem fly, *Melanagromyza sojae* (Diptera: Agromyzidae), on Sumatra: Seasonal incidence and the role of parasitism. *International Journal of Pest Management* 41: 127-133. DOI:10.1080/09670879509371937.

Venette RC. 2015. The challenge of modelling and mapping the future distribution and impact of invasive alien species. Report. [cited 05 September 2015]. Available from: <http://www.fs.fed.us/nrs/pubs/jrnl/2015/nrs_2015_venette_001.pdf>.

Wang CL. 1979. Occurrence and life-history of *Melanagromyza sojae* on soybean. *Journal of Agricultural Research of China* 28: 217-223.

Wang J, Gai J. 2001. Mixed inheritance model for resistance to agromyzid beanfly (*Melanagromyza sojae* Zehntner) in soybean. *Euphytica* 122: 9-18.

Yadav P, Banerjee S, Gupta MP, Yadav VK. 2015. Effect of weather factors on seasonal incidence of insect-pests of soybean. *Technofame: A Journal of Multidisciplinary Advance Research* 4: 46-51.

Yorinori JT, Paiva WM, Frederick RD, Costamilan LM, Bertagnolli PF, Hartman GE, Godoy CV, Nunes Jr J. 2005. Epidemics of Soybean Rust (*Phakopsora pachyrhizi*) in Brazil and Paraguay from 2001 to 2003. *Plant Disease* 89: 675-677.

Zalucki MP, Daglish G, Firempong S, Twine P. 1986. The Biology and Ecology of *Heliothis-Armigera* (Hubner) and *Heliothis-Punctigera* Wallengren (Lepidoptera, Noctuidae) in Australia - What Do We Know. *Australian Journal of Zoology* 34: 779-814.

Ziaee M. 2012. Oilseed Pests, Oilseeds. InTech, [cited 05 September 2015]. Available from: < http://www.intechopen.com/books/oilseeds/oilseed-pests>.

Ziska LH, Blumenthal DM, Runion GB, Hunt Jr ER, Soltero HD. 2011. Invasive species and climate change: an agronomic perspective. *Climatic Change* 105: 13-42.

# CHAPTER II

# COMPLETE MITOCHONDRIAL GENOMES OF THE SOYBEAN STEM FLY (SSF)

# *MELANAGROMYZA SOJAE* (DIPTERA: AGROMYZIDAE)

---

Adapted from:

Jonas André Arnemann, Tom Walsh, Karl Gordon, Hugh Brier, Jerson Vanderlei Carús Guedes, Wee Tek Tay. 2015. Complete mitochondrial genome of the soybean stem fly *Melanagromyza sojae* (Diptera: Agromyzidae). *Mitochondrial DNA* (Accepted 09-Sept-2015, In press).

## 2.1 Introduction

The mitochondrial DNA (mtDNA) is small, compacted and highly conserved circular molecule ubiquitous in all living organisms. The concept of predominantly maternally mode of inheritance (e.g., see Crozier 1990) for the mtDNA has increasingly been challenged from various empirical experimental findings in recent times (e.g., Nunes et al. 2013; Dokianakis and Ladoukakis 2014) where significant paternal leakage at population wide scales, as well as recombination and heteroplasmic events in diverse organisms (e.g., see reviews by Rubinoff et al. 2006; Krishnamurthy and Francis 2012) have been reported. Nevertheless, with integrated research angles such as incorporating ecological data (e.g., known distribution range, known hosts), behavioural characteristics (e.g., feeding behaviour and/or damage to crop host patterns), morphological characters, and molecular DNA sequence data (e.g., the mtDNA Cytochrome Oxidase subunit I (COI) gene, nuclear DNA loci), can greatly assist with identifying, discovering and categorising biodiversity and invasive species (Rubinoff et al. 2006; Tay et al. 2013; Mastrangelo et al. 2014).

The mtDNA of insects typically has length that ranged between 15 and 18Kb (Cameron 2014a). The mitochondriald DNA genomes (mitogenomes) of most insects (and indeed, in the great majority of bilateral symmetric animals) characterised to-date have been shown to contain 37 genes, of which 13 are protein coding genes (PCGs), two are ribosomal RNAs genes (rRNAs), 22 transfer RNA genes (tRNAs) (for translation of the mtDNA PCGs), and has a region controlling replication and transcription (also termed A+T rich region) (Moritz et al. 1987; Crozier et al. 1989, Crozier 1990; Cameron 2014b). In most animals mitochondrial genes are transcribed on both strands, and terms used include the ‘H’ (heavy) and ‘L’ (light) strands, a reference to differences in

G+T content between the two stands that arises due to their asymmetric replication (Reyes *et al.* 1998; Cameron 2014a). In most insects the majority strand corresponds to the H strand and the minority to the L (Cameron 2014a).

Genes within the mtDNA genome (e.g., COI, Cytochrome *b* (Cyt*b*), NADH dehydrogenase 4 (ND4), etc.) have been widely applied to systematic studies (i.e., as gene trees derived from mtDNA genes would enable effects of maternal gene flow patterns to be traced, Crozier 1990) and biological conservation purposes (e.g., for *Collembola* in the Canadian Arctic region, Hogg and Hebert (2004); soil invertebrate species in Canadian boreal forest, Römbke et al. (2006), to assist with species confirmation (e.g., *Helicoverpa armigera* (Behere et al. 2007, 2008; Tay et al. 2013)), understanding of gene flow patterns (e.g., in *Helicoverpa armigera* (Behere et al. 2007); in the fire ants *Solenopsis invicta* (Ascune et al. 2011)), and for inferring spatial structures (e.g., Tay et al. 1997; Chapter IV).

Despite the various shortfalls of applying solely the 'DNA barcoding' exercise in species identification as argued by, e.g., Rubinoff et al. (2006), the high copy number, ease and speed to which the mtDNA genome can be isolated, and genes such as the mtDNA COI and Cyt*b* can be readily sequence analysed due to the lack of introns (e.g., Crozier et al. 1989; Crozier 1990; Armstrong and Ball 2005; Cameron 2014a) have helped retained the popularity of applying these genes as markers of choice when assisting with species status confirmation. In most cases, the 5'- (N-terminal) region of the mtDNA COI gene of approximately 630bp is preferentially sequence characterised to assist with confirming species status of a wide range of animals including invasive aquatic (e.g., fish (Ward et al. 2009; Hubert et al. 2008); Crustacea (Costa et al. 2007)), and terrestrial animals (e.g., aphids (Foottit et al. 2008); *Spodoptera* species (Nagoshi

et al. 2011)) that were introduced into new environments. Applying the mtDNA COI gene to assist with rapid confirmation of global invasive pests (e.g., *Bemisia tabaci* cryptic species complex, De Barro et al. 2010); *Helicoverpa armigera* noctuid moths (Behere et al. 2007; Tay et al. 2013; Chapter IV); the tussock moth *Lymantria dispar* (Armstrong and Ball 2005; Ball and Armstrong 2006); the Mediterranean fruit flies *Ceratitis capitata* (Barr et al. 2012); Agromyzidae leafminer flies (Scheffer et al. 2006)) have all greatly benefit from having a relevant DNA sequence database.

High-value globally significant agricultural crops (e.g., soybean, cotton) are often targeted by diverse insect pest complexes, and the use of mtDNA gene markers such as the mtDNA COI and Cyt*b* gene has been applied to study various pests species complex of these high-value food and fiber crops. Pest species targeting soybean that have been characterised at the mtDNA COI gene region included the armyworm *Spodoptera* species in the US (Nagoshi et al. 2011); *Chloridea* (*Heliothis*) *virescens* in Brazil (Albernaz et al. 2012); the southern green stink bug *Nezara viridula* (Kavar et al. 2006); and in the Australian soybean moth (*Aproaerema simplixella*) (Buthelezi et al. 2012). Another significant pest in soybean fields is the soybean stem fly (SSF) *Melanagromyza sojae* (Zehntner), although there is currently no molecular genetic study on this agromizid fly species.

*M. sojae* is a highly damaging pest to soybean crops in Eastern Asia (e.g., China; Wang and Gai 2001), Southern Asia (e.g., India and Nepal; Thapa 2012), South East Asia (e.g., Indonesia; Van Den Berg et al. 1998), and Eastern Europe (e.g., Russia; Strakhova et al. 2013). In Australia, SSF has recently gained importance as an emerging soybean pest. Since its reported presence in the Lockey Valley (Queensland, Australia) as well as being noted as having the potential to become an agricultural insect pest (Shepard

et al. 1983), the first major outbreaks have since been reported to cause significant damage to Australia's subtropical coastal soybean crops (Brier and Moore 2013).

The SSF has the potential to be a significant biosecurity threat for northern Australia (Shepard et al. 1983), and also in North and South America such as in Brazil (Hirose and Moscardi 2012). Reports indicated that SSF can infest 100% of soybean plants and tunnelled up to 70% of stem length (Singh and Singh 1990; Singh and Singh 1992). SSF larvae damage the soybean stem negatively impacting plant growth and soybean yield (Talekar 1989). Until now there has been no information on the mitochondrial DNA genome and/or partial COI gene sequence for this species.

In this chapter, suspected *M. sojae* (SSF) specimens were first identified by combining larval feeding behaviour on host plants and morphological characters, prior to molecular characterisation of the insect's complete mtDNA genome, followed by developing PCR primers that specifically target partial mtDNA COI, ATP8/ATP6, and ND4 gene regions that may be applied to future population genetic studies. This study also reports the molecular characterisation and annotation of all 13 protein coding genes (PCGs), and enabled complete mitogenome sequence polymorphism rates for three *M. sojae* to be estimated. Using partial mtDNA COI sequences of a representative individual, a Maximum Likelihood (ML) phylogeny was also constructed to help infer the phylogenetic position of the suspected SSF fly, and to help determine if such a species had previously been sequence characterised at the mtDNA COI gene region.

## 2.2 Material and Methods

### 2.2.1 Morphological identification

Four Agromyzidae larvae sampled directly from individual soybean stalks (as the larva feeds on parenchyma cells within the pith cavity of the stem and in doing so creates a feeding tunnel, Van der Berg et al. 1998) from Cruz Alta (Rio Grande do Sul (RS), Brazil) were sent to Mr Hugh Brier, Senior Entomologist at the Department of Agriculture, Fisheries and Forestry Queensland (Australia), for visual verification of species identity. For species confirmation, larval morphological characters - notably their distinctive posterior spiracles (Dempewolf 2004) - was used, with the Brazilian samples also being directly compared with *M. sojae* samples previously collected from *Glycine max* host (collector: H. Brier; date: 26-March-2013) at Casino, New South Wales, Australia.

### 2.2.2 Molecular characterization of the complete mitochondrial DNA genome

The DNA extraction method used was the same as described in Chapter III. Due to the general absence of molecular data for Agromyzidae flies, we applied the next generation sequencing (NGS) method to characterize the mitochondrial DNA genomes of three randomly chosen individuals (Sample 09, Sample 11 and Sample 21, see Chapter III), from Brazil for the purpose of developing robust PCR markers such as for the mtDNA COI gene. We prepared individual gDNA libraries for the Brazilian samples for NGS as previously described in Walsh (2014), prior to running on the Illumina MiSeq sequencer. We assembled the three mitogenomes using Geneious® R8 (Biomatters, Auckland, NZ), and annotated using MITOS (Bernt et al. 2013) as described in Tay et al. (2014) and Piper et al. (2015), followed by manual fine-scale annotations (e.g., to identify putative stop codons) using Geneious® R8.

### 2.2.3 Intra-species nucleotide diversity in *M. sojae*

Estimates of the nucleotide diversity (π) of all protein coding genes (PCGs) between the three SSF individuals was carried out by aligning the full mitochondrial genomes using Geneious R8. The putative origin of replication which was also high in A+T-base compositions was excluded when estimating genome-wide nucleotide diversity levels and patterns. The program DnaSP v5.10.01 (Librado and Rozas 2009) was used to estimate the nucleotide diversity and to visualise the levels of nucleotide diversity of the aligned partial mitogenomes between the three flies by a sliding window analysis using default options for both window length (100) and step size of 25.

### 2.2.4 SSF mtDNA PCR markers development and PCR optimisation

For developing standardised mtDNA PCR markers for population genetic diversity survey of SSF, three regions of the mtDNA genome (i.e., mtDNA COI, ATP8/ATP6, ND4) were identified for this purpose. The template DNA sequence for which the SSF mtDNA primer pairs were developed from was based on the aligned consensus sequence of the three mitogenomes generated in this study. The mtDNA COI gene was chosen because of existing sequence database (e.g., from iBoL) for other Agromyzidae flies, and surveying the mtDNA COI gene region of SSF would allow comparison with exisiting Agromyzidae sequence database. For the two additional pairs of mtDNA PCR primers, these were developed to target the mtDNA regions with the highest nucleotide diversity, estimated based on the sliding window analysis described above. Primers were designed based on the criteria described in Tay et al.

(2008) (i.e., minimal primer-dimers formation and heteroduplex structures, minimal false priming sites to avoid non-specific PCR amplification), and >60˚C theoretical melting temperature ($T_m$) based on the calculation of [2(A+T)˚ + 4(G+C)˚]. Primers were designed using Oligo Primer Analysis Software Version 7.60 (Molecular Biology Insights, Inc., Cascade, USA).

PCR conditions for the mtDNA COI primers were optimized by gradient PCR (at 48˚C to 62˚C; 8 increments) on a BIORAD PCR machine (model C1000 Thermal Cycler), and amplicons visualized on a 1.5% 1x TAE agarose gel stained with GelRed™ (Biotium, Cat. # 41003). The final optimised and standardised PCR profile for the SSF-COI-F/R primers consisted of: 95°C for 5 minutes (one cycle), 30 seconds each of 95°C, 61°C and 72°C (34 cycles), followed by a final extension cycle of 72°C for 5 minutes. PCR amplicons were kept at 4˚C post-PCR and stored at -20˚C until needed. PCR amplification of individual DNA samples was carried out in a 25 µL total reaction volume that contained 25 ng of genomic DNA, 0.5 µM each forward and reverse primer, 0.2 mM of dNTP's, 1× Phusion HF Buffer (NEB), and 1.25 units of Phusion DNA polymerase (NEB).

Amplicons were purified using the QIAquick® PCR purification Kit (Qiagen) prior to being used as DNA template for Sanger sequencing reaction using the ABI BigDye® dideoxy chain termination sequencing system V3.1 (Applied Biosystems). Sequencing reaction and post sequencing reaction clean-up were as specified by the sequencing facility. Sequencing was carried out at the Australian National University Biomolecular Resource Facility (ANU BRF).

### 2.2.5 Inter-species nucleotide diversity in Agromyzidae

Estimates of mean inter-species nucleotide diversity between the Msoj-COI-02 haplotype (the haplotype shared by both Brazilian states and the Australian samples; Chapter III), and other related Agromyzidae species involved trimming all sequences to 555bp in the final dataset. Both the evolutionary divergence and the average inter-species evolutionary diversity were estimated using the Maximum Composite Likelihood model (Tamura et al. 2004), and included all codon positions (i.e., $1^{st}$ + $2^{nd}$ + $3^{rd}$), while gaps and missing data were excluded. Evolutionary genetic analyses were conducted in MEGA6 (Tamura et al. 2013).

### 2.2.6 Phylogeny of Agromyzidae based on partial mtCOI gene

Phylogenetic inference based on partial mtCOI genes included sequences of related Agromyzidae flies (named as *Melanagromyza* sp. in iBoL DNA database and GenBank) and *Fergusonina taylori* (as the out group) aligned to nucleotide positions from 61 to 666 of the *M. sojae* mitogenome (KT597923). The Maximum-Likelihood phylogeny softward PhyML 3.0 (Guindon et al. 2010) was used for the phylogenetic analysis of *Melanagromyza* species, and selecting the 'automatic model selection' option for optimising substitution model and 1,000 bootstrap replications for estimating node confidence.

## 2.3 Results

### 2.3.1 Species confirmation by morphology

Based on the larval morphological characters, notably the “distinctive posterior spiracles which have a blunt, somewhat atrophied central horn” (Dempewolf 2004), all four larvae sent to QDAF for identification were confirmed as *M. sojae* (Fig. 1).

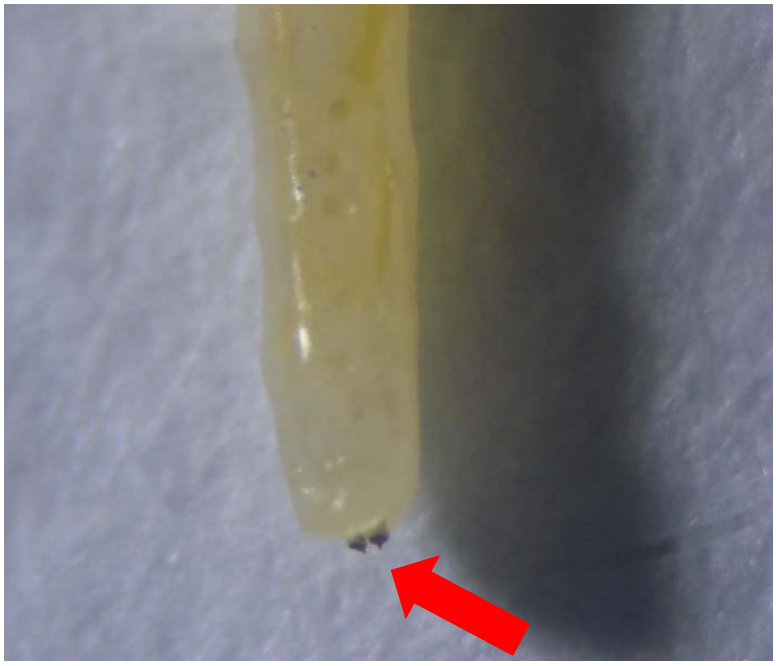

**Fig. 1:** Unique larval morphological character (indicated by the arrow) used for identification of *Melanagromyzae sojae* species. (Photo: Luis Eduardo Curioletti).

### 2.3.2 Molecular characterization of complete mitochondrial DNA genome

Sequence reads coverage (i.e., representing the number of times each base was sequence on average) from the Illumina MiSeq sequencing runs for *M. sojae* samples 09, 11, and 21 were 353x, 480x and 760x, respectively. The three *M. sojae* mitogenomes were independently assembled prior to mitogenome annotations. The complete (estimated) lengths of the three mitogenomes were 15,475bp (KT597923; Sample 11), 15,346bp (sample 21), and 15,173bp (sample 09). These length variations occurred at the predicted A+T-rich region within individual mitogenomes and reflected the difficulties and challenges of full mitogenome assembly via short reads from NGS methods as previously noted (Piper et al. 2015).

The presence of varying copy numbers of tandemly-repeated elements was reported as one of the characteristics of the insect A+T-rich region (Zhang and Hewitt 1997). Repetitive and low complexity DNA sequences (such as the A+T-rich region) are abundant in a broad range of species and are of considerable technical challenge for sequence alignment and assembly programs, and the use of next-generation sequencing with short read lengths and high data volumes, can significantly compound this bioinformatics challenge (Treangen and Salzberg 2012).

For all remaining regions of the partial mitogenomes between the three flies, no sequence length variation was detected. Results from the Mitos annotations of all three mitogenomes were identical and results of the sample 11 mitogenome annotation (with additional fine-scale manual annotations and Blastp search for confirmation) is presented in Table 1.

The additional two full mitochondrial genomes will be used to infer full motochindrial DNA phylogeny of 12 Agromyzidae species. The remaining Miseq data (e.g., SSF nuclear DNA, bacterial microbiota of SSF) will be used to develop alternative nuclear DNA markers, as microsatellite and EPIC-PCR DNA markers.

**Table 1:** Characteristics of the mitochondrial genome of *Melanagromyza sojae* with additional fine-scale manual annotations and Blastp search for confirmation. All PCG starts with the amino acid methionine (M) except ND5 which starts with Leucine (L). Ten of the PCGs also have the 'TAA' stop codon except for ND3, Cytb and ND1 which have 'TAG'.

| Gene* | Position | | Size Nucleotide (bp) | Codon | | | Intergenetic nucleotide | Strand |
|---|---|---|---|---|---|---|---|---|
| | From | To | | Amino acid | Start | Stop | | |
| **COI** | 1 | 1536 | 1536 | 511 | TTG (M) | TAA | 4 | H |
| trnL2 (taa) | 1541 | 1606 | 66 | | | | 2 | H |
| **CO2** | 1609 | 2391 | 783 | 240 | ATG (M) | TAA | -35 | H |
| trnK(ctt) | 2297 | 2366 | 70 | | | | 0 | H |
| trnD(gtc) | 2367 | 2432 | 66 | | | | 0 | H |
| **ATP8** | 2433 | 2591 | 159 | 52 | ATT (M) | TAA | -4 | H |
| **ATP6** | 2588 | 3262 | 675 | 224 | ATA (M) | TAA | 2 | H |
| **CO3** | 3265 | 4053 | 789 | 262 | ATG (M) | TAA | 7 | H |
| trnG(tcc) | 4061 | 4124 | 64 | | | | 15 | H |
| **ND3** | 4140 | 4478 | 339 | 112 | ATT (M) | TAG | -2 | H |
| trnA(tgc) | 4477 | 4539 | 63 | | | | 0 | H |
| trnR(tcg) | 4540 | 4601 | 62 | | | | 0 | H |
| trnN(gtt) | 4602 | 4667 | 66 | | | | 0 | H |
| trnS1(gct) | 4668 | 4735 | 68 | | | | 0 | H |
| trnE(ttc) | 4736 | 4801 | 66 | | | | 18 | H |
| trnF(gaa) | 4820 | 4886 | 67 | | | | -56 | L |
| **ND5** | 4831 | 6606 | 1776 | 591 | CTT (L) | TAA | -10 | L |
| **ND4** | 6597 | 8024 | 1428 | 475 | ATG (M) | TAA | -7 | L |
| trnH(gtg) | 6622 | 6685 | 64 | | | | ** | L |
| **nad4L** | 8018 | 8314 | 297 | 98 | ATG (M) | TAA | 2 | L |

| Gene* | Position | | Size Nucleotide (bp) | Codon | | | Intergenetic nucleotide | Strand |
|---|---|---|---|---|---|---|---|---|
| | From | To | | Amino acid | Start | Stop | | |
| trnT(tgt) | 8317 | 8381 | 65 | | | | 0 | H |
| trnP(tgg) | 8382 | 8447 | 66 | | | | 2 | L |
| **nad6** | 8450 | 8989 | 540 | 179 | ATT (M) | TAA | 3 | H |
| **Cytb** | 8993 | 10129 | 1137 | 378 | ATG (M) | TAG | -4 | H |
| **ND1** | 10126 | 11148 | 1023 | 340 | ATG (M) | TAG | 10 | H |
| trnS2(tga) | 10128 | 10193 | 66 | | | | *** | H |
| trnL1(tga) | 11159 | 11222 | 64 | | | | -42 | L |
| rrnL | 11181 | 12545 | 1365 | | | | -2 | L |
| trnV(tac) | 12544 | 12615 | 72 | | | | -1 | L |
| rrnS | 12615 | 13401 | 787 | | | | 0 | L |
| Replication origin | 13402 | 13980 | 579 | | | | 0 | ND |
| trnI(gat) | 13983 | 14047 | 65 | | | | 65 | H |
| trnQ(ttg) | 14113 | 14181 | 69 | | | | 3 | L |
| trnM(cat) | 14185 | 14253 | 69 | | | | 0 | H |
| **ND2** | 14254 | 15276 | 1023 | 340 | ATT (M) | TAA | -2 | H |
| trnW(tca) | 15275 | 15342 | 68 | | | | -8 | H |
| trnC(gca) | 15335 | 15400 | 66 | | | | 4 | L |
| trnY(gta) | 15405 | 15470 | 66 | | | | 5 | L |

*More details about the genes function can be found on Cameron (2014a,b); (** 'trnH(gtg)' is completely overllaping with the PCG ND4.; ***'trnS2(tga)' is completely overlapping with the PCG 'nad1'; ND: Not determined

The complete mitochondrial DNA genomes of the three *M. sojae* from Brazil all had 13 protein coding genes (PCGs), 2 ribosomal RNA (rRNA) genes and 22 tRNAs, with gene orientations as reported in other Agromyzidae flies (e.g., NC_015926; NC_016713; NC_016716; KR047789). Similar to other insects, the three mitogenomes have a A-T bias base composition (i.e., SAMPLE 11: 40.9% A, 36.7% T, 13.6% C, and 8.8% G; SAMPLE 21: 40.8% A, 36.7% T, 13.7% C, and 8.8% G; SAMPLE 09: 40.4% A, 36.6% T, 13.9% C, and 9.0% G). Characterisation of repeat units within the A+T-rich region involved a high degree of uncertainty due to contig assembly difficulties involving NGS short sequence reads, and these regions were therefore not annotated. A figure of the complete mitogenome of SAMPLE 11 showing the 13 PCGs, 22 tRNAs, 2 rRNAs, and the putative replication origin (AT-rich) region is presented in Fig.2.

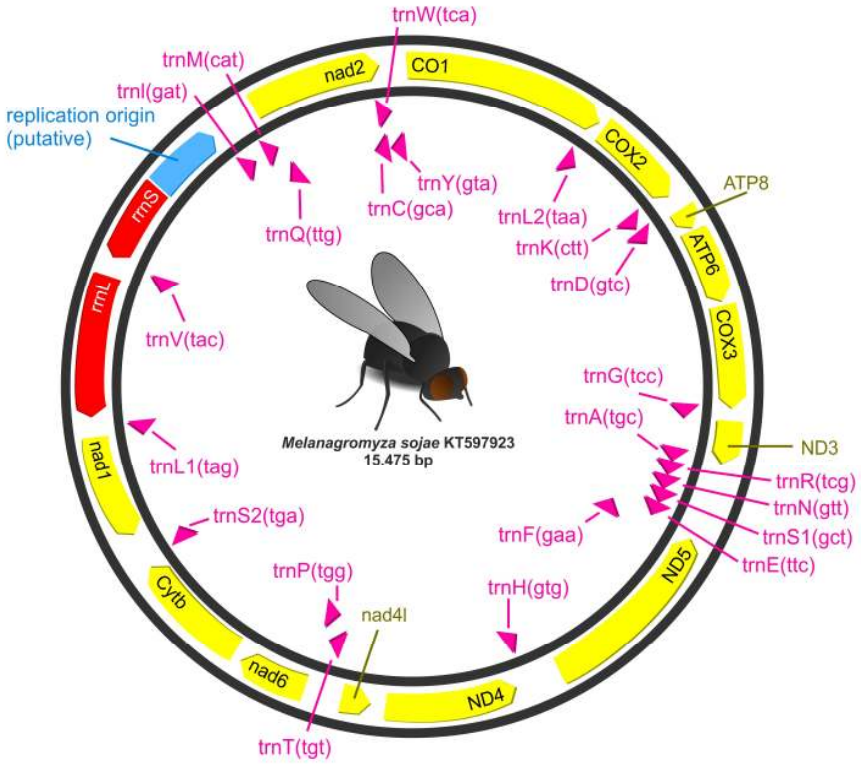


**Fig. 2:** The complete annotated mitogenome of the Agromizidae fly *Melanagromyza sojae* 'SAMPLE 11'. Directions of translation for 27 all genes within the mitogenome are indicated by arrows.

### 2.3.3 Intraspecies mitogenome PCG SNP patterns and nucleotide diversity

The total sequence length of all 37 genes excluding the putative replication origin in each mitogenome was 14,582bp and a total of 66 SNPs (64 transitions, 2 transversions) were detected in nucleotide comparisons between the three SSF mitogenomes (Table 2). A sliding windown analysis of nucleotide diversity (Pi) distribution patterns along the mitogenomes (excluding the AT-rich region) of the three Agromizidae flies were calculated (Fig. 3).

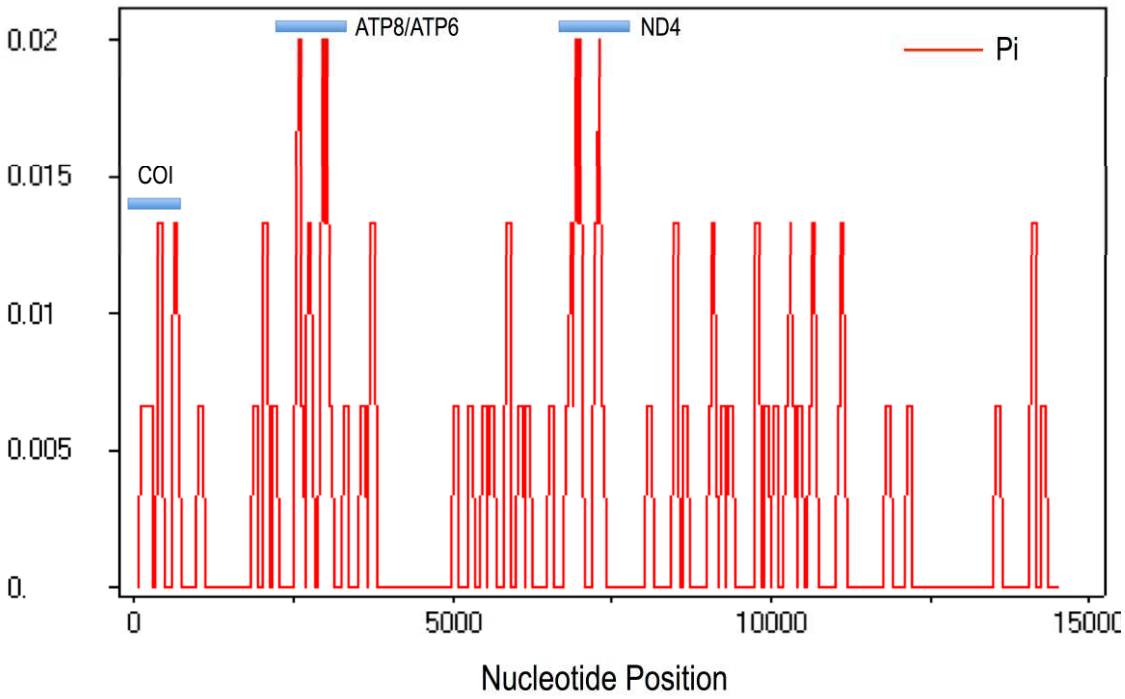


**Fig. 3:** Nucleotide diversity (Pi (π)) between three *M. sojae* flies from Brazil based on a sliding window analysis with window size of 100 and step size of 25. Bars indicate mtDNA gene regions (gene names provided) where PCR primers have been developed.

The nucleotide diversity (π) from the comparison of the three full mitogenomes (excluding the A+T-rich region) obtained was 0.00302 ± 0.0000008 s.e.

**Table 2:** Mitogenome SNP patterns from three *M. sojae* individuals (excluding the AT-rich region).

| PCG | COI | | | | | | | COII | | | | ATP8 | | ATP6 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **nt** | 144 | 237 | 388 | 414 | 648 | 675 | 1050 | 1884 | 2070 | 2074 | 2211 | 2572 | 2573 | 2602 | 2749 | 2773 | 2938 | 2977 | 2986 | 3058 |
| **SAMPLE 21** | C | G | T | G | T | C | A | G | A | G | G | C | T | T | T | A | G | G | A | A |
| **SAMPLE 11** | T | A | C | A | C | T | A | A | A | G | A | T | T | C | T | A | G | A | G | G |
| **SAMPLE 09** | T | A | C | A | C | T | G | A | G | A | G | C | C | T | C | G | A | A | G | G |

| COIII | | | | ND5 | | | | | | | | | ND4 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3309 | 3585 | 3732 | 3750 | 5038 | 5296 | 5479 | 5611 | 5859 | 5860 | 6064 | 6193 | 6538 | 6813 | 6888 | 6957 | 6969 | 7005 | 7257 | 7269 | 7344 |
| C | A | C | A | C | C | A | A | G | C | T | G | G | G | C | T | G | T | T | A | T |
| C | G | T | A | A | T | A | G | A | T | C | A | G | G | T | C | G | C | C | G | C |
| T | G | T | G | A | T | G | G | A | T | T | A | A | A | C | C | A | T | T | G | C |

| ND4L | ND6 | | | Cytb | | | | | | | | ND1 | | | | | | rRNA-L | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 8078 | 8492 | 8506 | 8674 | 9058 | 9107 | 9229 | 9370 | 9787 | 9790 | 9940 | 10061 | 10254 | 10347 | 10462 | 10645 | 10699 | 11086 | 11143 | 11843 | 12153 |
| G | T | T | C | G | A | A | A | T | C | C | G | C | C | C | G | C | T | T | C | C |
| A | T | T | T | A | G | A | G | C | T | C | A | C | T | C | A | T | C | C | T | T |
| A | C | C | T | A | G | G | A | C | T | A | A | T | T | T | A | C | C | C | T | T |

| ND2 | | | |
|---|---|---|---|
| 14453 | 15002 | 15017 | 15176 |
| C | C | A | C |
| C | A | G | C |
| T | A | A | T |

### 2.3.4 PCR primers design

A total of three PCR primer pairs were developed based on the consensus mtDNA genome sequence of the three *M. sojae* individuals. The primer pairs for the mtDNA COI gene region amplified an amplicon size of 906bp (mitogenome nucleotide positions (nt) 5 – nt911). For the remaining two primer pairs of ATP8/ATP6 and ND4, the expected amplicon sizes were 855bp (nt2514 – nt3519) and 633bp (nt6756 – nt7388), respectively. PCR profile for the mtDNA COI primer pairs, and also predicted PCR conditions for the untested ATP8/ATP6 and ND4 primer pairs, as well as primer sequences are provided in Table 3.

**Table 3:** PCR primer sequences for SSF, targeting the mtDNA COI 5' region (SSF-COI-F/R), and two high nucleotide diversity gene regions as identified based on a sliding window analysis (ATP8--ATP6; primers SSF-ATP8-F, SSF-ATP6-R), (ND4; primers SSF-ND4-F/R). PCR annealing temperature (˚C) has been optimised for the SSF-COI-F/R primers, and for the other two pairs were as predicted (*) from the primer designing program Oligo 7. Max (˚C) indicates the predicted maximum optimum annealing temperatures. N/A – not applicable.

| Primer | Sequence (5' – 3') | Gene(s) | Amplicon (bp) | ˚C | Max (˚C) |
|---|---|---|---|---|---|
| SSF-COI-F | GACAATGATTATTTTCGACAAAT | COI | 809 | 61 | N/A |
| SSF-COI-R | GTAAAATAAGCTCGTGTATCTACATC | | | | |
| SSF-ATP8-F | TATTATTGTTATTTTCCTTCTCTATC | ATP8-ATP6 | 855 | 48* | 60 |
| SSF-ATP6-R | CATTTAATTATTCCAGAAACAGTAG | | | | |
| SSF-ND4-F | CAAATATTCACGTAAATTACCTACA | ND4 | 633 | 48* | 61 |
| SSF-ND4-R | GTTTCAGGATCAATAATTTTAGC | | | | |

### 2.3.4 Inter-species diversity and maximum Likelihood phylogeny analysis of *M. sojae* and related Agromyzidae flies

A summary of known potential invasive species threats of Brazil (Hirose and Moscardi, 2012) and various significant Agromyzidae fly species to global agriculture based on published data are presented in Table 4.

**Table 4:** A list of Agromyzidae species with high dispersal ability including selected potential invasive species to Brazil.

| | Species | Region/country of occurrence | iBoL/GenBank Accession numbers | References |
|---|---|---|---|---|
| 1 | *Melanagromyza sojae* (Zehntner, 1900)* | Asia, Africa, Australia, Brazil | Yes, (BRA/AUS) | 1-3, this study |
| 2 | *Melanagromyza dolichostigma* (Meijere, 1922)* | Asia | No | 1-3 |
| 3 | *Melanagromyza pseudograta* (Spencer, 1977)* | Australia | No | 1 |
| 4 | *Melanagromyza shibatsuji* (Kato, 1961)* | Asia/Koreia, Japan | No | 1 |
| 5 | *Melanagromyza virens* (Loew, 1869) | North America/Canada, US | CNPPH748-12 | 9 |
| 6 | *Melanagromyza obtuse* (Malloch, 1909) | Asia | GBDP3918-07 | 2,8,10 |
| 7 | *Melanagromyza minimoides* (Spencer, 1966) | North America, Argentina, Bolivia, Venezuela, Guadeloupe | GBDP3919-07 (Bolivia) | 2 |
| 8 | *Melanagromyza cleomae* Spencer (1961) | Asia/Sri Lanka | GBDP3920-07 (Gannoruwa) | 8 |
| 9 | *Melanagromyza chalcosoma* (Spencer, 1959) | Africa | GBDP3921-07 (Kenya) | 2,8 |
| 10 | *Ophiomyia cornuta* (Meijere, 1910) | Australia | GBDP3915-07 | 5,7 |
| 11 | *Ophiomyia phaseoli* (Thyon, 1951)* | Asia, Africa, US, Europa, Australia, | GBDP3913-07 (Philippines) | 1-4 |

| | Species | Region/country of occurrence | iBoL/GenBank Accession numbers | References |
|---|---|---|---|---|
| 12 | *Ophiomyia spencerella* (Greatheat, 1969)* | Africa | No | 1-3 |
| 14 | *Ophiomyia centrosematis* (Meijere, 1940)* | Australia, Asia, Africa | No | 1,2 |
| 15 | *Ophiomyia quinta* (Spencer, 1969) | Canada, Europe | BBDED334-10 | 5,6 |
| 16 | *Ophiomyia lantanae* (Froggatt, 1919) | Australia, Sri Lanka, North and South America | GBDP3914-07 | 5,9 |

*indicate the species potentially 'invasive' to Brazil by (1). (1) Hirose and Moscardi (2012); (2) Dempewolf (2004); (3) Plantwise (2014); (4) EPPO (2014); (5) NCBI (2015); (6) Strakhova et al. (2013); (7) Scheffer et al. (2007); (8) IBOL (2015); (9) Fletcher (2003); (10) NPAG (2000).

Using partial Agromyzidae mtDNA COI gene sequences that are publicly available (i.e., *M. chalsosoma, M. minimoides, M. virens, M. obtuse, O. cornuta, O. lantanae, O. phaesoli and O. quinta* sequences; Table 4), the estimated mean inter-species evolutionary diversity between species is 13.2% (± 0.009 s.e.) based on 555bp of partial mtCOI gene in reported Agromizidae flies (including *M. sojae* haplotype Msoj-COI-02; i.e., the shared haplotype identified from both Brazilian states and the Australian samples, Chapter III).

The evolutionary divergence over sequence pairs against the *M. sojae* Msoj-COI-02 haplotype sequence were also estimated (Table 5). The evolutionary divergence between *M. sojae* and others species ranged from 0.140-0.190, with the most similar specie to *M. sojae* being *M. chalsosoma* (0.140), and the most divergent being *O. cornuta* (0.190).

**Table 5**: Estimates of evolutionary divergence over sequence pairs between *M. sojae* (Msoj-COI-02 haplotype, see Chapter III) from this study and related species. Standard error (s.e.) estimates from 500 bootstrap replications are shown in the upper triangle (in blue).

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | Msoj-COI-02 | | 0.013 | 0.014 | 0.015 | 0.014 | 0.015 | 0.013 | 0.014 | 0.014 | 0.014 |
| **2** | *M. chalsosoma* | 0.110 | | 0.014 | 0.014 | 0.014 | 0.014 | 0.014 | 0.014 | 0.014 | 0.014 |
| **3** | *M. obtusa* | 0.133 | 0.133 | | 0.015 | 0.014 | 0.015 | 0.016 | 0.014 | 0.015 | 0.015 |
| **4** | *O. cornuta* | 0.135 | 0.126 | 0.130 | | 0.014 | 0.014 | 0.012 | 0.012 | 0.013 | 0.014 |

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **5** | *M. minimoides* | 0.128 | 0.124 | 0.137 | 0.142 | | 0.015 | 0.015 | 0.015 | 0.013 | 0.014 |
| **6** | *M. cleomae* | 0.133 | 0.126 | 0.151 | 0.133 | 0.146 | | 0.015 | 0.015 | 0.014 | 0.014 |
| **7** | *O. lantanae* | 0.126 | 0.117 | 0.157 | 0.112 | 0.150 | 0.153 | | 0.012 | 0.013 | 0.014 |
| **8** | *O. phaesoli* | 0.132 | 0.135 | 0.133 | 0.108 | 0.148 | 0.153 | 0.115 | | 0.013 | 0.013 |
| **9** | *O. quinta* | 0.130 | 0.130 | 0.133 | 0.108 | 0.126 | 0.150 | 0.119 | 0.117 | | 0.014 |
| **10** | *M. virens* | 0.130 | 0.141 | 0.148 | 0.123 | 0.128 | 0.142 | 0.144 | 0.119 | 0.126 | |

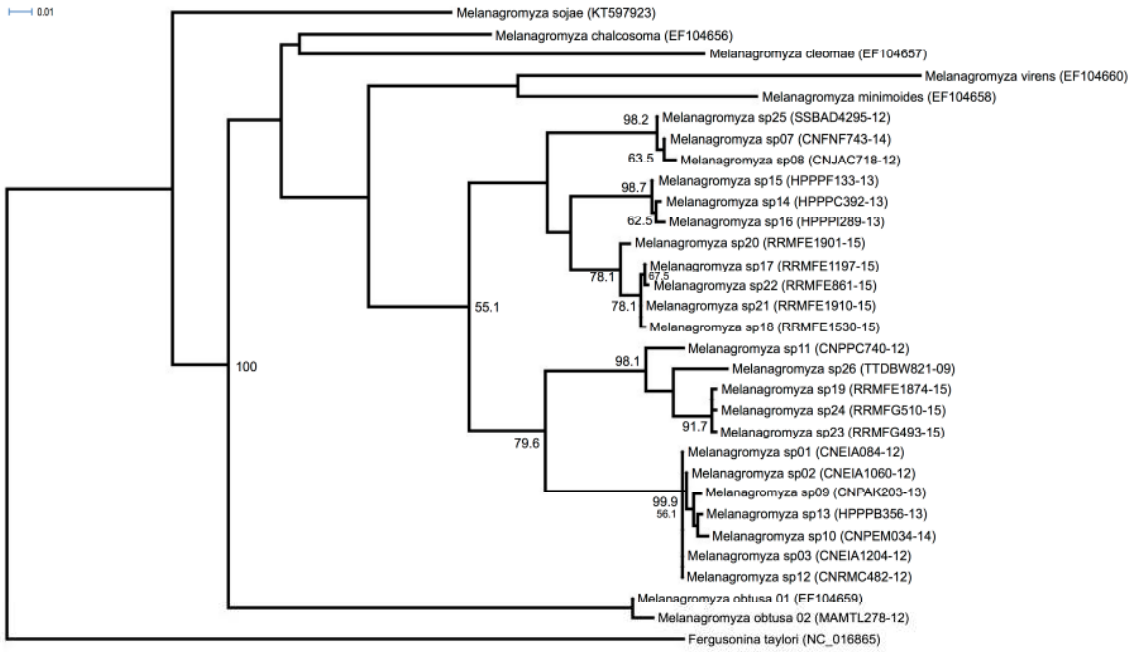


**Fig. 4:** Maximum Likelihood (ML) phylogeny analysis using PhyML (Guindon et al. 2010) of Agromyzidae species (substitution model: GTR+G6+I; 11086.49 (AIC); Gamma shape parameter: 0.597, proportion of invariable sites: 0.529), based on 607bp partial mtCOI gene. The out group is *Fergusonina taylori.* (NC_016865) (Scheffer et al. 2007). Bootstrap values of >50% are shown. The iBoL ID or GenBank accession numbers for all samples used are provided. All sequences used are aligned to nucleotide positions 61 to 666 of the *M. sojae* mitogenome (KT597923). Adapted from Arnemann et al. (2015, in press).

## 2.4 Discussion

The presence of *Melanagromyza sojae* in soybean fields in Brazil was first confirmed by larval morphology after suspected larvae were found damaging *G. max* crops in Cruz Alta in the state of Rio Grande do Sul (RS). This therefore enabled molecular characterisation of the pest's mitogenome to be carried out with confidence. For the purpose of informing relevant agricultural ministry authority (Dr. Decio Coutinho, Agricultural Defense Secretary) in a timely manner and to present both morphological evidence and molecular DNA confirmation, it was therefore also deemed necessary to ascertain sequence identity, with the mtDNA COI gene being the preferred genomic region due to pre-existing sequence data from other Agromyzidae flies. By applying the NGS platform, the complete mtDNA genome of this fly species was successfully determined, while also assisted with the development of a species-specific PCR primer pairs for the mtDNA COI gene region of *M. sojae*. PCR and subsequent sequencing of amplicons from samples collected from Brazil and *M. sojae* larvae collected (but poorly preserved) in 2013 in Australia showed that the mtDNA COI haplotype of the Australian *M. sojae* was also present in Brazil (Chapter III).

At last check when preparing this chapter (23 Aug 2015), no sequence data on *M. sojae* were available from public DNA database depositories (e.g., GenBank, iBoL). Results of the mitogenomes generated from this study therefore represent the first *M. sojae* mtDNA to be available for better understanding the population genetics and invasive biology of this pest fly species. The nucleotide diversity from comparing the three full mitogenomes was (0.00302 ± 0.0000008 s.e.), and is lower then the nucleotide diversity detected from partial mtDNA COI gene (630bp; 0.00408 ± 0.00047 s.e.) (Chapter III). Interestingly, these three complete mtDNA genomes all represented

unique haplotypes and contrasted that found when sequence surveyed only the 5' region of the mtDNA COI genes (see Chapter III). The mtDNA genetic diversity (and therefore the potential number of founding mothers) of *M. sojae* in Brazil will likely to be higher than detected based on single gene locus (Chapter III). Whether the high diversity of mtDNA genome haplotypes in *M. sojae* was in part associated with potential paternal mtDNA leakage as reported in another dipteran species (*D. melanogaster*; Nunes et al. 2013) will require further empirical experimental study.

The nucleotide diversity obtained through a sliding window analysis of the mitogenomes identified two genomic regions with highest nucleotide diversity (Fig. 3), potentially indicating genomic regions most informative for developing additional population genetic DNA markers for estimates of population-wide haplotype diversity (i.e., COI, ATP8/ATP6 and ND4; Table 3). Future studies should aim to determine the level of genetic diversity that can be achieved using these primers. The NGS data obtained may also allow specific nuclear genes (e.g., 28S rDNA, CAD) to be characterised, and therefore further bolster gene sets that will be available, and therefore contributing to future multigene phylogenetic studies such as the study of Scheffer et al. (2007) for inference of host use evolution within Agromyzidae species.

The inferred partial mtCOI phylogeny (Fig. 4) suggested a basal position for *M. sojae* as compared to other *Melanagromyza* species, while low bootstrap confidence (i.e., <50%) for various clades pointed to the need for further research on this agriculturally important and diverse dipteran group. New studies should consider markers proposed by this study but also other marker, e.g., 28S rDNA, (Dixon and Hillis 1993), CAD (a nuclear protein coding gene recently characterized for use as a phylogenetic marker in higher Diptera, Moulton and Wiegmann 2004), as well as the

full coding regions of the mitogenome. In Agromyzidae, reconstruction of larval feeding mode suggested that the ancestral feeding mode was leaf-mining rather than stem- or cambium-mining (Dempewolf 2005). The inclusion of multigene loci (e.g., COI, 28S rDNA and CAD markers) will allow comparisons between *M. sojae* with other recent and detailed Agromyzidae systematic studies, making possible to make inferences on host-use evolution, such as the ability to feeding on diverse plant families (Scheffer et al. 2007).

Applying NGS platform for rapid retrieval of genomic data from invasive pests species can lead to better understanding of pest incursion history (e.g., Staats et al. 2012) and test ecological hypotheses about dispersal patterns in invasive species (Chown et al. 2015; Richardson et al. 2011; e.g., inferring genetic structure/gene flow patterns in Brazilian populations of *Grapholita molesta*, Silva-Brandão et al. 2015), and should be adapted for understanding the population and evolutionary genetics of SSF in Brazil. This study which utilised NGS platform has resulted in rapid authoritative and informative communication with Brazilian Agricultural ministry to benefit agricultural sectors, while NGS data generated will be further analysed to better understand ecological consequences from the introduction of this specie. This study represents an example of how monitoring of invasive pests such as SSF using genetic tools is translating into the development of targeted management strategy for this species, while early detection of an invasion will likely have a positive flow-on effect such as cheaper control costs, reduced commodity loss (Hulme 2006; Rius et al. 2015), and minimise potential international trade embargos by trading partner countries.

## 2.5 References


Albernaz KC, Silva-Brandão KL, Fresia P, Cônsoli FL, Omoto C. 2012. Genetic variability and demographic history of *Heliothis virescens* (Lepidoptera: Noctuidae) populations from Brazil inferred by mtDNA sequences. *Bulletin of Entomological Research* 102: 333-343.

Armstrong K. 2010. DNA barcoding: a new module in New Zealand's plant biosecurity diagnostic toolbox. *OEPP/ EPPO Bulletin* 40: 91-100.

Armstrong KF, Ball SL. 2005. DNA barcodes for biosecurity: invasive species identification. *Philosophical Transactions of the Royal Society B* 360: 1813-1823.

Barr NB, Islam MS, DeMeyer M, McPheron BA. 2012. Molecular Identification of *Ceratitis capitata* (Diptera: Tephritidae) using DNA Sequences of the COI Barcode Region. *Annals of the Entomological Society of America* 105: 339-350.

Brier, H, Moore N. 2013. Soybean stem fly outbreak in soybean crops. The Beatsheet. [cited 05 September 2015]. Available from: <http://thebeatsheet.com.au/crops/pulses/soybeans/soybean-stem-fly-outbreak-in-soybean-crops/>.

Buthelezi NM, Conlong DE, Zharare GE. 2012. The groundnut leaf miner collected from South Africa is identified by mtDNA CO1 gene analysis as the Australian soybean moth (*Aproaerema simplixella*) (Walker) (Lepidoptera: Gelechiidae). *African Journal of Agricultural Research* 7: 5285-5292.

Cameron SL. 2014a. How to sequence and annotate insect mitochondrial genomes for systematic and comparative genomics research. *Systematic Entomology* 39: 400-411. doi: 10.1111/syen.12071.

Cameron SL. 2014b. Insect Mitochondrial Genomics: Implications for Evolution and Phylogeny. *Annual Review of Entomology* 59: 95-117.

Chown SL, Hodgins KA, Griffin PC, Oakeshott JG, Byrne M, Hoffmann AA. 2015. Biological invasions, climate change and genomics. *Evolutionary Applications* 8: 23–46.

Costa FO, Waard JR, Boutillier J, Ratnasingham S, Dooh RT, Hajibabaei M, Hebert PDN. 2007. Biological identifications through DNA barcodes: the case of the Crustacea. *Canadian Journal of Fisheries and Aquatic Sciences* 64: 272-295.

Crozier RH, Crozier YC, Mackinlay AG. 1989. The CO-I and CO-II region of honeybee mitochondrial DNA: evidence for variation in insect mitochondrial evolutionary rates. *Molecular Biology and Evolution* 6: 399-411.

Crozier RH. 1990. From population genetics to phylogeny: uses and limits of mitochondrial DNA. *Australian Systematic Botany* 3: 111-124.

Dempewolf M. 2005. Dipteran leaf miners. In: Raman, A., Schaefer, C.W., Withers, T.M. (Eds.), Biology. In: Ecology, and Evolution of Gall-inducing Arthropods, vol. 1. Science Publishers Inc., Plymouth, UK, pp. 407-429.

Dixon, MT, Hillis DM. 1993. Ribosomal RNA secondary structure: Compensatory mutations and implications for phylogenetic analysis. *Molecular Biology and Evolution* 10: 256-267.

Dokianakis E, Ladoukakis ED. 2014. Different degree of paternal mtDNA leakage between male and female progeny in interspecific *Drosophila* crosses. *Ecology and Evolution* 4: 2633-2641.

Ficetola GF, Miaud C, Pompanon F, Taberlet P. 2008. Species detection using environmental DNA from water samples. *Biology Letters* 4: 423-425.

Foottit G, Maw HEL, Von Dohlen CD, Hebert, PDN. 2008. Species identification of aphids (Insecta: Hemiptera: Aphididae) through DNA barcodes. *Molecular Ecology Resources* 8: 1189-1201.

Groenenberg DSJ, Pirovano W, Gittenberger E, Schilthuizen M. 2012. The complete mitogenome of *Cylindrus obtusus* (Helicidae, Ariantinae) using Illumina next generation sequencing. *BMC Genomics* 13: 114.

Guidolin AS, Fresia P, Consoli FL. 2014. The genetic structure of an invasive pest, the Asian Citrus Psyllid *Diaphorina citri* (Hemiptera: Liviidae). *PLoS ONE* 9(12): e115749. doi:10.1371/journal.pone.0115749.

Hebert PDN, Cywinska A, Ball SL, Waard JR. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B* 270: 313-321.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. 2004. Identification of birds through DNA barcodes. *PLoS Biology* 201-206.

Hirose E, Moscardi F. 2012. Insetos de outras regiões do mundo. IN: EMBRAPA-Soja: manejo integrado de insetos e outros artrópodes-praga. Brasília, DF: Embrapa, 859 p.

Hogg ID and Hebert PDN. 2004. Biological identification of springtails (Hexapoda: Collembola) from the Canadian Arctic, using mitochondrial DNA barcodes. *Canadian Journal of Zoology* 82:749-754.

Hubert N, Hanner R, Holm E, Mandrak NE, Taylor E, Burridge M, et al. 2008. Identifying Canadian Freshwater Fishes through DNA Barcodes. *PLoS ONE* 3(6): e2490. doi:10.1371/journal.pone.0002490.

Hulme PE. 2006. Beyond control: Wider implications for the management of biological invasions. *Journal of Applied Ecology* 43: 835:847.

Kavar T, Pavlovcic P, Susnik S, Meglic V, Virant-Doberlet M. 2006. Genetic differentiation of geographically separated populations of the southern green stink bug *Nezara viridula* (Hemiptera: Pentatomidae). *Bulletin of Entomological Research* 96: 117-128.

Kircher M, Kelso J. 2010. High-throughput DNA sequencing: concepts and limitations. *Bioessays* 32: 524-536.

Librado P, Rozas J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452.

Mahon AR, Nathan LR, Jerde CL. 2014. Meta-genomic surveillance of invasive species in the bait trade. *Conservation Genetics Resources* 6: 563-567.

Moritz C, Dowling TE, Brown WM. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Review of Ecology and Systematics* 18: 269-292.

Moulton JK, Wiegmann BM. 2004. Evolution and phylogenetic utility of CAD (rudimentary) among Mesozoicaged Eremoneuran Diptera (Insecta). *Molecular Phylogenetics and Evolution* 31: 363-378.

Nagoshi RN, Brambila J, Meagher RL. 2011. Use of DNA Barcodes to Identify Invasive Armyworm *Spodoptera* Species in Florida. *Journal of Insect Science* 11: 154.

Nunes MD, Dolezal M, Schlotterer C. 2013. Extensive paternal mtDNA leakage in natural populations of *Drosophila melanogaster*. *Molecular Ecology* 22: 2106-2117.

Rius M, Bourne S, Hornsby HG, Chapman MA. 2015. Applications of next-generation sequencing to the study of biological invasions. *Current Zoology* 61: 488-504.

Römbke J, Jänsch S, Scroggins R. 2006. Identification of potential organisms of relevance to Canadian boreal forest and northern lands for testing of contaminated soils. *Environmental Reviews* 14: 137-167.

Rosa AJM, Sonoda KC. 2010. Aplicação de DNA barcoding no levantamento de espécies de insetos. Planaltina, DF: Embrapa Cerrados, 2010.

[cited 22 August 2015]. Available from: <http://www.cpac.embrapa.br/noticias/artigosmidia/publicados/217/>.

Rubinoff D. 2006. Utility of mithochondrial DNA barcodes in species conservation. *Conservation Biology* 4: 1026-1033.

Scheffer SJ, Lewis ML, Joshi RC. 2006. DNA Barcoding Applied to Invasive Leafminers (Diptera: Agromyzidae) in the Philippines. *Annals of the Entomological Society of America* 99: 204-210.

Shepard, M. Lawn RJ. Schneider MA. 1983. Insects on Grain Legumes in Northern Australia: A Survey of Potential Pests and Their Enemies. University of Queensland Press. Australia. 81pp.

Silva-Brandão KL, Silva OABN e, Brandão MM, Omoto C, Sperling FAH. 2015. Genotyping-by-sequencing approach indicates geographic distance as the main factor affecting genetic structure and gene flow in Brazilian populations of *Grapholita molesta* (Lepidoptera, Tortricidae). *Evolutionary Applications* 8(5):476-485. doi:10.1111/eva.12257.

Smith MA, Fisher BL, Hebert PDN. 2005. DNA barcoding for effective biodiversity assessment of a hyper diverse arthropod group: the ants of Madagascar. *Philosophical Transactions of the Royal Society B: Biological* 360: 1825-1834.

Spencer KA. 1990. Host specialization in the world Agromyzidae (Diptera). Series Ent. Kluwer Academic Publishers, Dordrecht. 45: 1-444.

Staats M, Erkens HJJ, van de Vossenberg B, Wieringa JJ, Kraaijeveld K, Stielow B, Geml J, Richardson JE, Bakker FT. Genomic Treasure Troves. 2013. Complete genome sequencing of herbarium and insect museum specimens. *PLoS One* 8(7): e69189.

Tay WT, Beckett SJ, De Barro PJ. 2014. Phosphine resistance in Australian *Cryptolestes* species (Coleoptera: Laemophloeidae): Perspective from mitochondrial DNA Cytochorme Oxidase I analysis. *Pest Management Science*. doi: 10.1002/ps.3805. [Epub ahead of print].

Treangen TJ and Salzberg SL. 2012. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature Reviews Genetics* 13: 36-46 doi:10.1038/nrg3117.

Van den Berg H. Shepard BM, Nasikin. 1998. Response of soybean to attack by stemfly *Melanagromyza sojae* in farmers' fields in Indonesia. *Journal of Applied Ecology* 35: 514-522.

Ward RD, Hanner R, Hebert PDN. 2009. The campaign to DNA barcode all fishes, FISH-BOL. *Journal of Fish Biology* 74: 329-356.

Zhang DX, Hewitt GM. 1997. Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochemical Systematics and Ecology* 25: 99-120.

# CHAPTER III

# GENETIC DIVERSITY OF *MELANAGROMYZA SOJAE* (DIPTERA: AGROMYZIDAE) IN SOYBEAN (*GLYCINE MAX*) IN BRAZIL

---

Adapted from:

Jonas André Arnemann, Wee Tek Tay, Tom Walsh, Hugh Brier, Karl Gordon, Frederico Hickmann, Jerson Vanderlei Carús Guedes. 2015. The Soybean Stem Fly *Melanagromyza sojae* (Diptera: Agromyzidae) in the New World: detection of high genetic diversity from soybean fields in Brazil. *Journal of Pest Science* (Submitted: 27-Sept-2015).

## 3.1 Introduction

Soybean (*Glycine max*) is one of the most important vegetable crops worldwide with high levels of international trade, and the largest harvested area of any global commodity (Macdonald et al. 2015). Soybean has its centre of origin and domestication in north-eastern Asia (i.e., China and adjacent regions; Chung and Singh 2008), and is currently grown on all continents except the Arctic and Antarctic. In Brazil, soybean is cultivated in over 31.5 million hectares (CONAB 2015), and over 53 million hectares across the South American continent (FAOSTAT 2015).

Invasive pest species are exotic organisms that are able to proliferate and persist in the environment where they invade (Mack et al. 2000). The arrival of invasive species into a new region are increasingly associated with human activities such as tourism (e.g., *Anthonomus grandis* in Brazil, suggested to have arrived from the United States aboard air traffic, rather than by natural expansion; Ramalho and Santos 1994), increased movements of goods between countries (e.g., *Tuta absoluta* introduced into Brazil in 1979 from another country from South America; Michereff Filho and Vilela 2001), human migration (*cf*. trade movements) (e.g., distribution of the Yellow Crazy Ant *Anoplolepis gracilipes*; Lowe et al. 2001) and/or growth of international trade activities (e.g., the spider mite *Schizotetranychus hindustanicus* from India was reported in Brazil in 2010, Navia and Marsaro Jr. 2010; Capinera 2008). Arthropods and plant pathogens can also enter into new environments via marketed agricultural, horticultural, and/or livestock products (De Barro et al. 2011; Campbell 2001), although intentional introduction of animals, such as for recreational purposes (e.g., rabbits and foxes into Australia from the UK, Flux 1994; Short and Smith 1994), to serve as biological control agents (e.g., cane toads (*Bufo marinus*) in Australia,

introduced to control sugar cane beetle species; Van Beurden 1981) are excellent examples of invasive pests with significant negative environmental and socio-economic impact.

Biological incursions by alien/non-endemic species can also be by natural dispersal (i.e. dispersal without human intervention). For example, the relocation of large numbers of desert locusts, *Schistocerca gregaria* (Forskall), to the Caribbean and northern South America from Africa in 1988 is hypothesised to have been caused by weather (Capinera 2008). Recently, physiological adaptation underpinned by genetic selection has also been shown to be an important factor in natural establishment of alien species in novel environments, as reported for the long distance migratory behaviour of the monarch butterfly *Danaus plexippus*, whereby migratory North American populations' global dispersal was predominantly due to selection on flight muscle function (Zhan et al. 2014). In the European Union member countries, significant economic losses to the livestock industry in recent times have been, in parts, results of climate change that promoted long distance dispersals via northward expansion of the bluetongue virus (e.g., VTB8) vector species *Culicoides imicoloa* (Purse et al. 2006).

Relationship between trading countries in a commodity that may potentially be infested by a quarantine pest (e.g., an insect) can be hindered by the lack of a standardized quarantine protocol to assist with rapid and unambiguous identification of intercepted suspect insect species. With increasing international trade, invasive pest species are introduced, often unintentionally (e.g., global dispersal of the red imported fire ants *Solenopsis invicta* through global trade activities, Ascunce et al. 2001). If such alien species are introduced to places where there are no natural enemies, a suitable

climate and abundant accessible resources, these exotic organisms may become a significant issue, (i.e., Enemy Release Hypothesis (Keane and Crawley 2002), but this hypothesis remains controversial due to difficulty of testing (Jeschke et al. 2012)). Certain biological characteristics predispose insects to establish successfully in new environments, and these may include factors such as: (i) small body sizes such that it can more easily escape detection, (ii) having good flight capability, (iii) high fecundity/reproduction rates, and (iv) are pre-adapted to the new environment (Capinera 2008) and (v) polyphagy.

In Brazil the Environmental Crimes Law (Art. 61 of the Federal Law No. 9.605/98) was enacted to regulate the introduction of animals/pests into the country. This Environmental Crimes Law regards the spread of diseases, pests, or species that may cause damage to agriculture, livestock, fauna, flora or ecosystems as environmental crimes (Leão et al. 2011). Brazilian agriculture has had a long history of encountering issues of exotic pest incursions that significantly affected agricultural production, with increasing frequencies of invasive agricultural pest incursions being observed since the start of 1900's (Lopes-da-Silva et al. 2014). The recent introduction and spread of the boll weevil (*Anthonomus grandis*) from North/Central America in Brazil is a good example (Reaser et al. 2005). *A. grandis* was first detected in 1983 in the states of São Paulo and Paraíba, from where it spread to most producing areas at incredible speed. Within 10 years after its first detection, *A. grandis* had successfully established across all Brazilian cotton producing states (Lukefahr et al. 1994). Other recent invaders included the invasive whitefly *Bemisia tabaci* MEAM 1 (previously known as 'B biotype') (De Barro et al. 2011), the Asian Soybean Rust (*Phakosphora*

*pachyrhizi*) (Yorinori et al. 2005), and the Old World cotton bollworm (*Helicoverpa armigera*) (Czepak et al. 2013; Tay et al. 2013; Lopes-da-Silva et al. 2014).

Another group of invasive insect pests is the agriculturally important agromyzids flies, usually intercepted at country borders during import plant quarantine inspections (e.g., interceptions since 1940's at the USA country borders; and at the Japanese and European borders; Kamiji and Iwaizumi 2013; Roque and Auger-Rozenberg 2006; USDA 1941). The first record of Agromyzidae flies infesting soybean fields in Brazil was in the state of Rio Grande do Sul in 1983, in the Municipalities of Passo Fundo and Santa Maria (Gassen and Schneider 1985). Link et al. (2009) subsequently published a new report that described the occurrence of agromyzids flies in five soybean fields at the Municipality of São Francisco de Assis (Fig. 1-A). Whether these two incidents represented separate incursion events, or if populations in the Municipality of São Francisco de Assis were the progenies of the Rio Grande do Sul populations has not been investigated.

Morphological identification between Agromyzidae species is achieved using the shape of the male genitalia, via larval morphological characters, and aided by larval feeding habits (Thapa 2012; Dempewolf 2004). However, morphological species identification typically requires extensive training and taxonomic knowledge, and is especially so in view of worldwide Agromyzidae fly species diversity (e.g., >2000 species of agromizid flies reported for cultivated and wild host plants; Thapa 2012). The soybean stem fly (SSF; also known as the soybean stem miner (SSM)) *Melanagromyza sojae* (Zehntner) belong to the family Agromyzidae and is highly polyphagous (i.e., attacking plants such as *Glyicne max, Phaseolus vulgaris, Pisum sativum, Vigna angularis,* and other members from the Fabacea Family; Dempewolf

2004). *M. sojae* has been reported in diverse global regions (Dempewolf 2004) excluding North and South Americas, and is regarded as one of the most important pests in soybean fields in parts of Russia (Strakhova et al. 2013), in Asia (e.g., China (Wang and Gai 2001); India and Nepal (Thapa 2012)), and in parts of South East Asia (e.g., Indonesia (Van Den Berg et al. 2008)) and is a potential pest for Brazil (Hirose and Moscardi 2012) and for northern Australia (Shepard et al. 1983).

Identification of *M. sojae* has relied on adult and/or larval morphological characters, such as through the defining larval antrophied posterior spiracles morphology specific to the SSF (H. Brier, Queensland Department of Agriculture and Fisheries; pers. comm.; see Chapter II (Fig. 1)). In this study, survey results of Brazilian *M. sojae* population genetic diversity are presented using the newly developed SSF-specific mtCOI molecular marker (Chapter II), and potential implications of *M. sojae* incursion in Brazil will be discussed.

## 3.2 Material and methods

### 3.2.1 Samples

Larvae and adults of suspected *M. sojae* were collected from soybean plants in fields from Rio Grande do Sul (RS) and Santa Catarina (SC) states, Brazil, in the season of 2014/15 (Table 1; Fig. 1). At individual sampling sites, fly larvae were collect from soybean plants that showed feeding damages, and reared in the lab to obtain adults prior to preserving them in individual 1.5mL Eppendorf tubes with 1000 µL of 99.9% ethanol for molecular characterization purposes. Two *M. sojae* samples previously collected by Mr H. Brier (Senior Entomologist, Queensland Department of Agriculture and Fisheries, Queensland, Australia) on 26-March-2013 from *G. max* host at Casino,

New South Wales, Australia, were also included to enable direct comparison of the mitochondrial DNA cytochorome oxidase I (COI) gene. In addition, four Agromyzidae larvae sampled directly from individual soybean stalks from Cruz Alta (Rio Grande do Sul (RS), Brazil) were also sent to Mr. Hugh Brier for morphological verification of species identity.

**Table 1:** Collection sites and dates, mtDNA COI GenBank accession numbers of *Melanogromyza sojae* specimens from Brazil and Australia. Brazilian states are Rio Grande do Sul (RS) and Santa Catarian (SC). The Australian samples were from Casino in the state of New South Wales (NSW). All *M. sojae* specimens were collected from soybean (*Glycine max*) host.

| Sample ID | City | State | Sampling date | GenBank Accession number |
|---|---|---|---|---|
| 1 | Boa Vista do Buricá | RS | 18-April-2015 | KT821473 |
| 2 | Boa Vista do Buricá | RS | 18-April-2015 | KT821495 |
| 3 | Campo Novo | RS | 18-April-2015 | KT821481 |
| 4 | Cruz Alta | RS | 18-April-2015 | KT821474 |
| 5 | Cruz Alta | RS | 18-April-2015 | KT821475 |
| 6 | Cruz Alta | RS | 18-April-2015 | KT821493 |
| 7 | Novo Machado | RS | 17-April-2015 | KT821489 |
| 8 | Três de Maio | RS | 17-April-2015 | KT821486 |
| 9 | Três de Maio | RS | 17-April-2015 | KT821490 |
| 10 | Tucunduva | RS | 17-April-2015 | KT821476 |
| 11 | Tucunduva | RS | 17-April-2015 | KT821477 |
| 12 | Descanso | SC | 15-April-2015 | KT821483 |
| 13 | Descanso | SC | 15-April-2015 | KT821494 |
| 14 | Iporã do Oeste | SC | 16-April-2015 | KT821478 |
| 15 | Iporã do Oeste | SC | 16-April-2015 | KT821482 |
| 16 | Iporã do Oeste | SC | 16-April-2015 | KT821487 |
| 17 | Mondaí | SC | 15-April-2015 | KT821496 |
| 18 | Mondaí | SC | 15-April-2015 | KT821492 |
| 19 | Riqueza | SC | 15-April-2015 | KT821488 |
| 20 | Riqueza | SC | 15-April-2015 | KT821497 |
| 21 | Tunapólis | SC | 15-April-2015 | KT821479 |
| 22 | Tunapólis | SC | 15-April-2015 | KT821480 |
| 23 | Tunapólis | SC | 15-April-2015 | KT821491 |
| 24 | Casino | NSW, AU | 26-March-2013 | KT821484 |
| 25 | Casino | NSW, AU | 26-March-2013 | KT821485 |

### 3.2.2 Total genomic DNA (gDNA) extraction

A total of 23 suspected Brazilian *M. sojae* specimens (Table 1) underwent molecular characterization at the Commonwealth Scientific and Industrial Research Organization (CSIRO) Black Mountain Laboratories in Canberra, Australia. Individual specimens were washed three times in 3x 1000µL 99.9% ethanol prior to gDNA extraction. With the exception of three suspect adult *M. sojae* flies where non-destructive gDNA extraction method (Tay et al. 2014) was used, total gDNA from all remaining specimens was extracted from either legs or partial thorax (for adults) or whole larval body using Qiagen® DNasy Blood and Tissue DNA Extraction Kit (Cat. # 69506). Final elution volume for individual gDNA sample was in 35µL of Qiagen buffer EB (Cat. # 19086), with gDNA quality ascertained by 1.5% agarose gel visualization, and concentration quantified using Qubit® 2.0 Fluorometer (Invitrogen™).

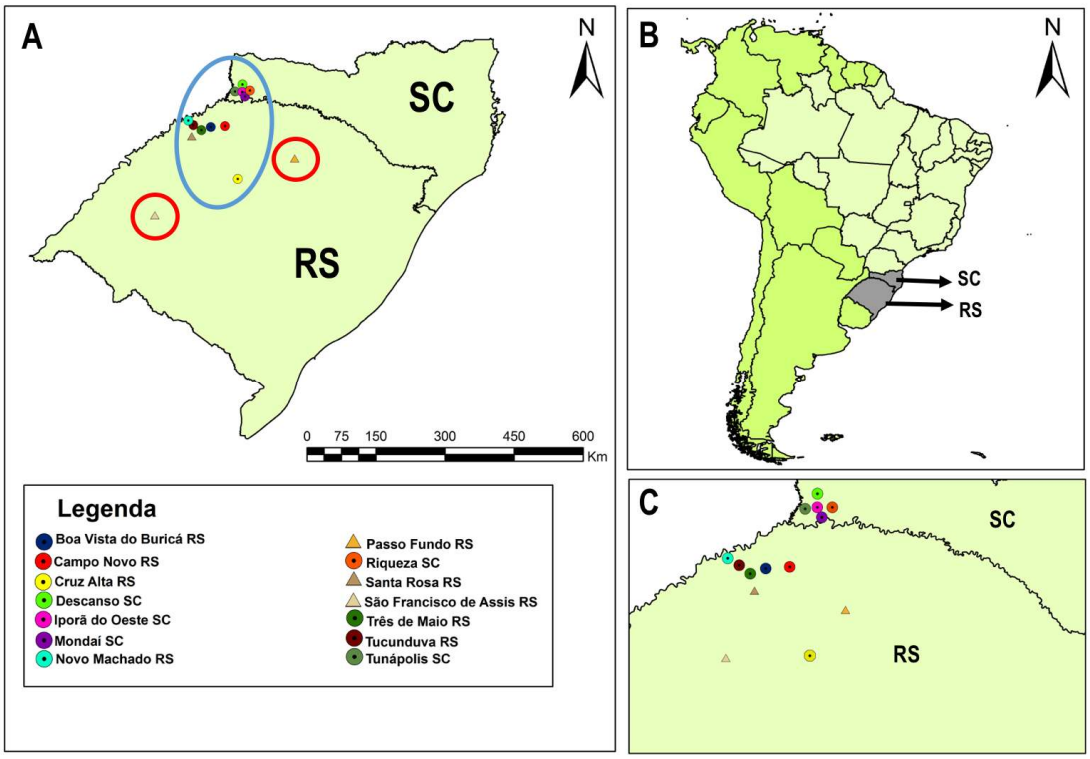


**Figure 1:** A map of sampling sites in Brazil from which suspected *M. sojae* were collected (blue circle in 'A'). The red circles in 'A' indicates the locations where individuals of *Melanagromyza* sp. were found in 1983 and 2009 but were not identified at the species level. Brazilian states are Santa Catarina (SC) and Rio Grande do Sul (RS).

### 3.2.3 Primers, PCR amplification and sequencing of partial mitochondrial DNA COI gene

Primers to amplify a partial mtDNA COI gene region (Chapter II) were designed based on the mitochondrial DNA genome (mitogenome, KT597923) of a *M. sojae* fly (Arnemann et al. 2015) using the primer design and analysis software Oligo 6 (Molecular Biology Insights, Inc., DBA Oligo, Inc.) (see Chapter II). Primers were designed to amplify the mtDNA COI region so as to be comparable to other Agromyzidae flies in the iBoL and NCBI GenBank databases. PCR conditions were optimized on a BIORAD PCR machine (model C1000 Thermal Cycler), and amplicons visualized on a 1.5% 1x TAE agarose gel stained with GelRed™ (Biotium, Cat. # 41003).

A 906bp fragment of the mtDNA COI gene was PCR amplified using the primers SSF-COI-F01 (5' GACAATGATTATTTTCGACAAAT 3') and SSF-COI-R01 (5' GTAAAATAAGCTCGTGTATCTACATC 3') (Chapter II) using the following PCR profile: 95°C for 5 minutes (one cycle), 30 seconds each of 95°C, 61°C and 72°C (34 cycles), followed by a final extension cycle of 72°C for 5 minutes. PCR amplicons were incubated at 4˚C post-PCR and kept at -20˚C until needed. PCR amplification of individual DNA samples was carried out in a 25 µL total reaction volume that contained 25 ng of genomic DNA, 0.5 µM each forward and reverse primer, 0.2 mM of dNTP's, 1× Phusion HF Buffer (NEB), and 1.25 units of Phusion DNA polymerase (NEB).

Amplicons were purified using the QIAquick® PCR purification Kit (Qiagen) prior to being used as DNA template for Sanger sequencing reaction using the ABI BigDye® dideoxy chain termination sequencing system V3.1 (Applied Biosystems). Sequencing reaction and post sequencing reaction clean-up were as specified by the sequencing facility. Sequencing was carried out at the Australian National University Biomolecular Resource Facility (ANU BRF).

### 3.2.4 Sequence analysis and molecular characterization of mtDNA COI gene

The programs Pregap and Gap4 within the Staden package (Staden et al. 2000) were used for editing and analyzing DNA sequences and to generate sequence contigs. Assembled partial mtDNA COI contigs were checked for premature stop codons that may indicate a pseudogene using Geneious® R8 (Biomatters Ltd., New Zealand) and by Blastp. Sequences that differed by one or more nucleotides were considered as different haplotypes, while sequences exhibiting identical SNPs at same nucleotide positions were considered as a same haplotype.

### 3.2.5 MtDNA haplotypes network and COI phylogenetic analysis

A mtDNA COI haplotype network for *M. sojae* was constructed manually and verified using the statistical parsimony phylogenetic network estimation program TCS v1.21 (Templeton et al. 1992). The distribution map of the SSF haplotypes was built using PopArt <http://popart.otago.ac.nz>.

Estimates of evolutionary divergence between all *M. sojae* individuals (n=25) was across a *ca.* 630bp region of the 5' (N-terminal) end of the mtDNA COI gene. Evolutionary genetic analyses between *M. sojae* haplotypes were conducted in MEGA6 (Tamura et al. 2013). Estimates of haplotype diversity ((h) ± S.E.) and nucleotide diversity ((π) ± S.E.) were carried out using the molecular evolution software package DNA Sequence Polymorphism (DnaSP) version 5.10.01 (Librado and Rozas 2009).

## 3.3 Results

### 3.3.1 Species confirmation

Based on the larval morphological characters, notably the "distinctive posterior spiracles which have a blunt, somewhat atrophied central horn" (Dempewolf 2004), all four larvae sent to QDAF for identification were confirmed as *M. sojae* (Fig. 1, Chapter II), although molecular confirmation via the partial mtCOI gene was not carried out on these specimens due to time constrains. The partial mtDNA COI gene region of specimens that originated from Brazil were analysed against the mtDNA sequences of two confirmed *M. sojae* specimens from Casino (NSW), Australia. This comparison identified a mtCOI haplotype (i.e., Msoj-COI-02; Table 2) to be 100% identical between three Brazilian flies (n = 1 from RS; n = 2 from SC) and the two Australian *M. sojae* samples, adding support to the morphological analysis that *M. sojae* is in Brazil. The

three *M. sojae* adults that were non-destructively extracted for their gDNA were deposited in the Australian National Insect Collection (ANIC) at CSIRO, Canberra (ANIC Database numbers 29035952, 29035953 and 29035954).

### 3.3.2 PCR amplification and sequence analysis

Of the 906bp mtDNA partial COI region amplified using the primers developed specifically for SSF, single nucleotide polymorphisms (SNPs) were absent at nucleotide positions (nt) 631 to nt 906 in all Brazilian (n = 23) and Australian (n = 2) SSF individuals. This region was therefore trimmed for all *M. sojae* data set, leaving 630bp as for sequences from the iBoL DNA database. The low estimates of evolutionary divergence between all *M. sojae* sequences were as expected at the intra-species level (e.g., Scheffer 2000) and ranged from 0-0.01% (± 0.001-0.004 S.E.).

### 3.3.3 Haplotypes patterns

A total of 10 haplotypes (GenBank accession numbers KT821473-97) were identified from 23 individuals collected from Santa Catarina (SC) and Rio Grande do Sul (RS) states of Brazil, and two individuals from New South Wales, Australia (Table 2). From the trimmed 630bp of partial mtCOI gene, 12 base substitutions were identified and all involved transition (i.e., purine/purine; pyrimidine/pyrimidine) substitutions (nine T ✶ C; three G ✶ A; Table 2). Based on the SSF population-wide base substitution patterns a ‘consensus’ SNP profile was determined (Table 2) that also matched the Msoj-COI-02 haplotype.

**Table 2**: Single nucleotide polymorphism (SNP) and haplotypes identified in field-collected *Melanagromyza sojae* samples from soybean crops in Brazil's Santa Catarina (SC) and Rio Grande do Sul (RS) states, and from Casino, New South Wales (NSW), Australia. Consensus single nucleotide polymorphisms (SNP's) (= Msoj-COI-02 haploytpe) are determined by population majority, nucleotide changes identical to the consensus are indicated by '.'.

| **SNP nucleotide position** | | **9** | **64** | **81** | **99** | **192** | **231** | **309** | **315** | **343** | **369** | **603** | **630** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Consensus** | | **T** | **G** | **C** | **T** | **G** | **C** | **T** | **T** | **C** | **A** | **T** | **T** | |
| **Sample ID** | **State** | | | | | | | | | | | | | **Haplotype** |
| 1 | RS | . | . | . | . | A | . | . | . | . | . | C | . | Msoj-COI-01 |
| 4 | RS | . | . | . | . | A | . | . | . | . | . | C | . | Msoj-COI-01 |
| 5 | RS | . | . | . | . | A | . | . | . | . | . | C | . | Msoj-COI-01 |
| 10 | RS | . | . | . | . | A | . | . | . | . | . | C | . | Msoj-COI-01 |
| 14 | SC | . | . | . | . | A | . | . | . | . | . | C | . | Msoj-COI-01 |
| 21 | SC | . | . | . | . | A | . | . | . | . | . | C | . | Msoj-COI-01 |
| 22 | SC | . | . | . | . | A | . | . | . | . | . | C | . | Msoj-COI-01 |
| 11 | RS | . | . | . | . | A | . | . | . | . | . | C | . | Msoj-COI-01 |
| 3 | RS | . | . | . | . | . | . | . | . | . | . | . | . | Msoj-COI-02 |
| 15 | SC | . | . | . | . | . | . | . | . | . | . | . | . | Msoj-COI-02 |
| 12 | SC | . | . | . | . | . | . | . | . | . | . | . | . | Msoj-COI-02 |
| 24 | NSW | . | . | . | . | . | . | . | . | . | . | . | . | Msoj-COI-02 |
| 25 | NSW | . | . | . | . | . | . | . | . | . | . | . | . | Msoj-COI-02 |
| 8 | RS | . | . | T | . | . | . | . | . | . | . | . | C | Msoj-COI-03 |
| 16 | SC | . | . | T | . | . | . | . | . | . | . | . | C | Msoj-COI-03 |
| 19 | SC | . | . | T | . | . | . | . | . | . | . | . | C | Msoj-COI-03 |
| 7 | RS | . | . | . | C | . | . | . | . | T | G | . | C | Msoj-COI-04 |
| 9 | RS | . | . | . | C | . | . | . | . | T | G | . | C | Msoj-COI-04 |
| 23 | SC | . | . | . | C | . | . | . | . | T | G | . | C | Msoj-COI-04 |
| 18 | SC | . | A | . | C | . | . | . | . | T | G | . | C | Msoj-COI-05 |
| 6 | RS | . | . | . | . | . | T | . | . | . | . | . | . | Msoj-COI-06 |
| 13 | SC | . | . | . | | . | . | . | C | . | . | . | . | Msoj-COI-07 |
| 2 | RS | . | . | . | . | A | . | C | . | . | G | C | . | Msoj-COI-08 |
| 17 | SC | . | . | . | C | . | . | . | . | T | . | . | . | Msoj-COI-09 |
| 20 | SC | C | . | . | . | A | . | . | . | . | . | . | . | Msoj-COI-10 |

The base substitutions within this haplotype are representative of majority SNPs across all detected haplotypes and are shared between Brazil and Australian populations (albeit at a lower frequency possibly due to low sampling size). This suggests that the Msoj-COI-02 haplotype is likely to be the ancestral haplotype, although accurate inference of phylogenetic relationship between haplotypes will require analysis of multigene from the mtDNA genome or longer mtDNA COI gene region (Winkler et al. 2009). Unique and shared haplotypes at the national level (i.e., between SC and RS states) and between countries (Brazil and Australia) were also ascertained, with unique haplotypes within Brazil constituted 50% of all haplotypes identified in the limited individuals sampled to-date (Table 3).

**Table 3:** Number of unique and shared haplotypes identified in different states of Brazil and Australia. Number of individuals (n) examined are listed. States within Brazil are Santa Catarina (SC), and Rio Grande do Sul (RS), and New South Wales (NSW) in Australia.

| Country | State (individuals sampled) | Number of haplotypes | Number of unique haplotypes | Number of shared haplotypes |
|---|---|---|---|---|
| **Brazil** | SC (n = 12) | 8 | 4 | 4 |
| **Brazil** | RS (n = 11) | 6 | 2 | 4 |
| **Brazil** | SC + RS (n = 23) | 10 | 5 | 4 |
| **Australia** | NSW (n = 2) | 1 | 0 | 1 |

The highest SSF nucleotide and haplotype diversity was found in Santa Catarina State. The pairwise uncorrected (p) genetic distances between all *M. sojae* haplotypes were low (ranged: 0 - 0.01) as expected for the intra-species level comparison (Table 5).

**Table 4**: Comparison of *M. sojae* mtDNA partial COI nucleotide diversity ($\pi$) and haplotype diversity (h) between different Brazilian states.

| Local | Nucleotide diversity | Haplotype diversity |
|---|---|---|
| **Santa Catarina State (SC)** | 0.00446 ± 0.00064 | 0.924 ± 0.057 |
| **Rio Grande do Sul State (RS)** | 0.00428 ± 0.00081 | 0.800 ± 0.114 |
| **Brazil (SC+RS)** | 0.00408 ± 0.00047 | 0.853 ± 0.048 |

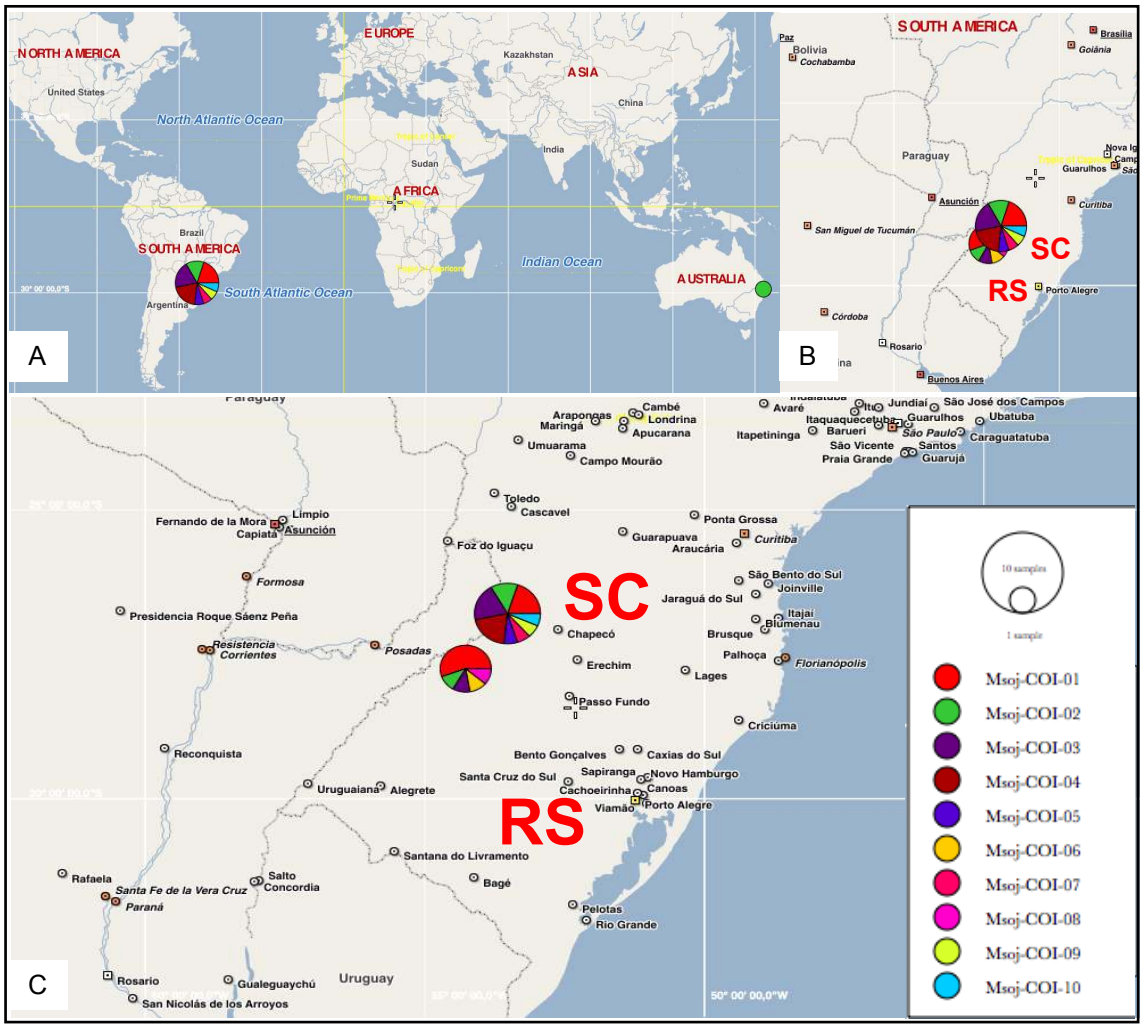


**Figure 2**: The haplotype distribution patterns and diversity of *Melanagromyza sojae* in: Australia and Brazil (panel A); and enlarged map detailing *M. sojae* collection sites in the states of Santa Catarina (SC), and Rio Grande do Sul (RS) in Southern Brazil (panels B and C).

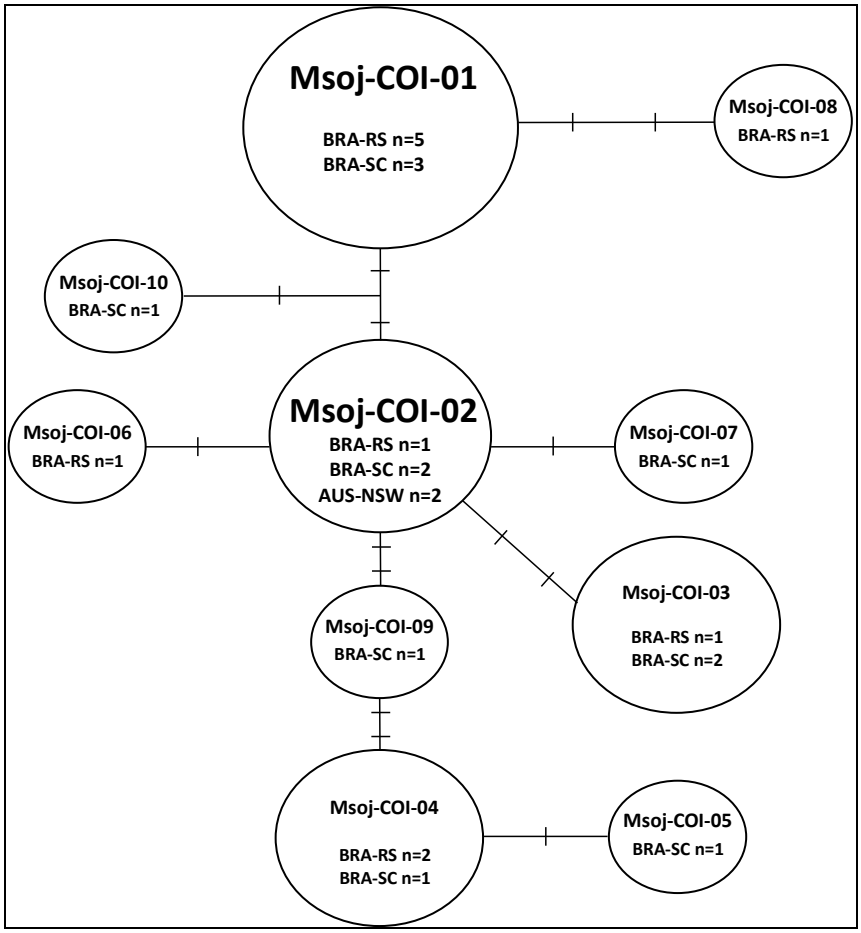


**Figure 3**. Haplotype network of *M. sojae* based on partial (740 bp) mtDNA COI gene, and included samples from Australia (AUS-NSW) and Brazil (BRA-RS, BRA-SC). Each haplotype is represented by a circle, and is identified by 'Msoj-COI-01' to 'Msoj-COI-10'. Haplotype 'Msoj-COI-01' included 8 individuals; haplotype 'Msoj-COI-02', 'Msoj-COI-03' and 'Msoj-COI-04' have 5, 3 and 3 individuals respectively. All remaining haplotypes have 1 individual each. Number of base changes that differentiated between haplotypes are represented by a black bar.

**Table 5**: Estimates of evolutionary divergence (nucleotide distances) between all *M. sojae* haplotypes based on 630bp of partial mtCOI gene. Standard error (s.e.) estimates from 500 bootstrap replications are shown in the upper triangle (in blue).

| | Msoj-COI-01 | Msoj-COI-02 | Msoj-COI-03 | Msoj-COI-04 | Msoj-COI-05 | Msoj-COI-06 | Msoj-COI-07 | Msoj-COI-08 | Msoj-COI-09 | Msoj-COI-10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Msoj-COI-01 | | 0.002 | 0.002 | 0.003 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 |
| Msoj-COI-02 | 0.003 | | 0.002 | 0.003 | 0.003 | 0.001 | 0.001 | 0.003 | 0.002 | 0.002 |
| Msoj-COI-03 | 0.005 | 0.003 | | 0.003 | 0.003 | 0.002 | 0.002 | 0.003 | 0.003 | 0.002 |
| Msoj-COI-04 | 0.008 | 0.005 | 0.005 | | 0.001 | 0.003 | 0.003 | 0.003 | 0.002 | 0.003 |
| Msoj-COI-05 | 0.010 | 0.007 | 0.007 | 0.001 | | 0.003 | 0.003 | 0.003 | 0.002 | 0.004 |
| Msoj-COI-06 | 0.004 | 0.001 | 0.004 | 0.007 | 0.008 | | 0.002 | 0.003 | 0.002 | 0.002 |
| Msoj-COI-07 | 0.004 | 0.001 | 0.004 | 0.007 | 0.008 | 0.003 | | 0.003 | 0.002 | 0.002 |
| Msoj-COI-08 | 0.003 | 0.005 | 0.008 | 0.008 | 0.010 | 0.007 | 0.007 | | 0.003 | 0.003 |
| Msoj-COI-09 | 0.005 | 0.003 | 0.005 | 0.003 | 0.004 | 0.004 | 0.004 | 0.008 | | 0.003 |
| Msoj-COI-10 | 0.003 | 0.003 | 0.005 | 0.008 | 0.010 | 0.004 | 0.004 | 0.005 | 0.005 | |

### 3.4 Discussion

The nucleotide diversity ($\pi$, i.e., degree of intra population polymorphism) of Brazilian *M. sojae* populations was 0.0045 and 0.0043, while the haplotype diversities (h, i.e., the uniqueness of particular haplotypes in the population) were 0.924 and 0.800 for Santa Catarina and Rio Grande do Sul, respectively. To put this in context, the nucleotide diversity was higher than various dipterans' genetic diversity studies (e.g., *Drosophila lacertosa* lineages in China, with $\pi$=0.0018 to 0.00324, He et al. 2007; *Liriomyza sativae* ($\pi$=0.00068 to 0.00300) in China, Du et al. 2014). The SSF diversity is similar with other invasive pests in South America, e.g., the lepidopteran *H. armigera* (Chapter IV; Mastrangelo et al. 2014; Leite et al. 2014). The number of haplotypes identified in the relatively small sample sizes within Brazil suggests significant genetic diversity in populations of *M. sojae*, and is reminiscent of the invasive genetic signature of the Old World cotton bollworm *H. armigera* genetic diversity detected in Brazil to-date (e.g., Tay et al. 2013; Mastrangelo et al. 2014; Leite et al. 2014; Chapter IV).

The SSF is a polyphagous species feeding on soybean plants as its main host, as well as on other cultivated legume crops (Verma et al. 1989). Findings from both Chapter II and this chapter confirmed *M. sojae* to be present in the soybean agricultural landscapes of both SC and RS states, possibly with established populations. Together with the substantial soybean area available in neighbouring Uruguay (1.3 million ha; FAOSTAT, 2015) and Argentina (20 million ha; FAOSTAT, 2015; which is the 3rd largest world producer of soybean, Bolsa de Cereales 2015), there is a significant likelihood that *M. sojae* can establish and sustain population growth in South America. As all SSF samples were taken from soybean plants in the same season, it was not

possible to correlate haplotypes with either host plants or seasons. Follow-up studies, to better understand the ecology of SSF in novel environments, should include detailed studies on host use, as well as gene flow patterns within Brazilian states (e.g., Parana, Santa Catarina, Rio Grande do Sul) and between neighbouring countries (e.g., Paraguay, Uruguay, Argentina) over multiple seasons, such as has been reported for *H. armigera* (Leite et al. 2014).

The mitochondrial DNA marker developed has enabled an estimate of the minimal number of maternal lineages in the Brazilian SSF populations. However, to obtain a more complete picture of the history and evolutionary potential of invasive populations and for inferring in-depth population-genetic structure and dynamics, effective nuclear DNA markers must also be developed as part of the standard suite of molecular genetic investigative tools (Zhang and Hewitt, 2003). These studies will be essential for future identification of potential population origin(s), to assist with the search for potential biological control agents and to help develop effective management tools, e.g., to enable genotypic matching of soybean varieties that may be resistant to SSF (Wang and Gai 2001). Furthermore, information on distribution patterns, genetic diversity and population structure will ensure the implementation of effective SSF management policies. For example, understanding the gene flow patterns based on more DNA markers and larger sampling sizes of SSF will help ascertain whether SSF populations within Brazil can be treated as a panmictic population (e.g., as found for *Triatoma infestans*, a vector of the causative agent of Chagas disease in Bolivia, Giordano et al. 2005), which would enable the simplification and unification of management strategies.

This study follows the recent confirmation of *M. sojae* in Brazil (Chapter II), reporting the unexpectedly high genotype diversity in populations of *M. sojae* from soybean fields in two southern Brazilian states. Previously, Gassen and Schneider (1985) reported the occurrence of *Melanogromyza* sp. in Brazil, but the species was not determined. Interestingly, the *Melanagromyza* sp. occurrence reported by Gassen and Schneider (1985) was in the Municipality of Passo Fundo (RS state), only 100 kilometres from some current sampling sites. Where circumstances permit (e.g., voucher samples were located and deemed of sufficiently good quality for molecular genetic analysis purposes), future studies should revisit historical samples (i.e. specimens collected in 1983 and 2009 by Gassen and Schneider (1985) and Link et al. (2009), respectively) to enable better understanding of the history of SSF incursion into South America.

Recent advances in molecular genetic tools have greatly improved the possibility of ascertaining the most likely scenario(s) (i.e. natural, accidental, and/or deliberate introduction) under which invasive pests arrived in the New World. Specifically, museums and quarantine/boarder protection agencies (e.g., 'Carlos Ritter' Natural Sciences Museum in Rio Grande do Sul, and 'Dr. Fritz Plaumann' Entomology Museum in Santa Catarina) together with International Agricultural Defence System (VIGIAGRO), the DSV (National Plant Protection Organization-NPPO in Brazil), and the System of Management Information for International Traffic Products and Agricultural Input (SIGVIG) (BRASIL 2012), are important biological resource centres for forensic investigation to determine incursion pathways and frequencies of exotic organisms.

Increasingly the value of historical samples (especially relating to global invasive pests) is being realised. For example, Davies et al. (1999) was able to reconstruct the

invasion history of the Mediterranean fruit fly *Ceratitis capitata* in the USA via sequence variation at four intron loci. The authors showed that individuals captured in the introduced range across different years represented separate introduction events, rather than as an original infestation that had persisted at low levels. Similarly, Tay et al. (2012) was able to characterise the partial mtCOI gene region of Gennadius' original 1889 whitefly specimen, deposited at the Smithsonian National Museum of Natural History, to confirm the true identity of the invasive *B. tabaci* as the 'Mediterranean (MED)' species within the *Bemisia* pest cryptic species complex. Interpreting potential incursion pathways and inferring population origins such as from intercepted specimens at quarantine facilities/national ports should also consider the likely complexity associated with transport network (e.g., Keller et al. 2011), where the source population of an exotic organism may not be the last port of call due to high network connectivity of international trade.

The global shipping network presents risks when invasive species use ports as 'stepping-stones' in a sequence of invasions in which an invader spreads from its endemic area to other places that become the origins for further spread (Keller et al. 2011). Brazil has a high risk for the introduction and establishment of exotic insects because of its continental dimensions, large border region, diversed climate (equatorial, tropical and temperate) (Nimer 1979; IBGE 2015), and extensive agricultural trade with various countries (Mata and Freitas 2008; Oliveira et al. 2012). These factors are likely to have played significant roles in recent high profile biosecurity incidents (e.g., Chapter IV; the rice water weevil *Oryzophagus oryzae*, Costa Lima 1936; *Thrips palmi*, Monteiro et al. 1995; the oriental fruit moth *Grapholita molesta*, Silva et al. 1962).

Spencer (1973) described the plant genus *Glycine* as a host of the following species of Agromyzidae flies: *Tropicomya vigneae* (Seg.), *Ophiomyia phaseoli* (Tryon), *O. shibatsuji* (Kato), *O. centrosematis* (de Meijere), *Japanagromyza tristella* (Thomson), *Melanagromyza dolichostigma* (de Meijere), *M. koizumii* (Kato), *M. vignalis* (Spencer), and *M. sojae* (Zehntner) (restricted to Asia and Africa). Elsewhere, *M. sojae* has been reported to be present in India, China, Japan, South Korea, Taiwan, Vietnam, Laos, Philippines, Indonesia, Malaysia, Saudi Arabia, Australia, Solomon Islands, Micronesia, Israel, Egypt, and South Africa (Dempewolf 2004). Due to the likely presence of cryptic Agromizidae species and possible misidentifications, genetic differentiation is required (e.g., Scheffer 2000). Despite surveys confirming the presence of *M. sojae* in these diverse global regions, genetic diversity such as based on the maternal mtDNA COI marker at the population level has not been conducted. Findings from this study will serve as the basis for a more comprehensive mtDNA database (e.g., Chapter IV) for future comparisons with both native populations endemic to specific world regions, as well as to populations representing invasive ranges, thereby contributing to better understanding of potential historical and future global incursion pathways of *M. sojae*.

Preliminary observations made in other countries (e.g., India, Nepal and Indonesia; Thapa 2012; Van Den Berg et al. 1995) suggest that *M. sojae* is capable of becoming a serious threat to many legume crops in the southern part (i.e. the 'Cone Sul' region) of South America. Follow up studies must necessarily involve monitoring the spread of *M. sojae* in Brazilian soybean fields, providing information relating to potential origin(s) and developing management strategies for this pest. Whilst mtDNA marker have been effective in assisting with the rapid identification of *M. sojae* in Brazil, genomic data generated from NGS technology (e.g., Chapter II) should be

applied to better understand the metagenomics of this agriculturally and economically important pest fly species, as well as for mining and developing candidate nuclear markers (e.g., microsatellite DNA and/or EPIC-PCR markers) for monitoring of gene flow patterns and to infer population structure at local (national) and global scales.

Many host plants of *M. sojae* are cultivated in South America (Dempewolf 2004) and represent economically important crops of the 'Cone Sul' region. The high reproduction rate and capacity for natural spread (e.g., in Iran, Ziaee 2012) probably render containment or eradication expensive and unachievable. In addition, early infestations are difficult to detect (small flies and oviposition scars). For the moment, data is lacking on the potential of establishment of *M. sojae* in the Cone Sul region, but the fact that *M. sojae* have been detected across diverse eco-climatic zones at a global scale (Dempewolf 2004) would suggest that it has the potential to also establish across the North and South America continents, and its potential distributional range may benefit from pre-emptive population modelling to assist with biosecurity preparedness, as has been carried out for *H. armigera* (Kriticos et al. 2015).

In Brazil the Bt soybean varieties used express Cry1Ac proteins and its effect on SSF is unknown. The recent paper from Yu et al. (2014) mentioned the presence of detectable levels of Cry1Ac toxin in the adult stage of SSF collected from a Bt soybean field, but did not mention possible effects on the SSF development. Furthermore, the main Bt toxins that are known to affect dipteran insects are Cry4Aa, Cry4Ba, Cry11Aa, Cyt1Aa, Cry10Aa and Cyt2Ba (Ben-Dov 2014). Data relating to life history, evolutionary genetics, ecology, and parameters associated with integrated pest management strategies and chemical responses that are necessary to better understand and manage SSF remained lacking. Future investigations in southern Brazil will need to

focus on these aspects, such as seed treatment control and potential damage, chemical response to Bt soybean (the preferred GM-technology in Brazil agriculture practices), effectiveness of biological control (e.g., by parasitoids, Ven Den Berg et al. 1995; Shepard and Barrion 1998) including species from potential native range of *M. sojae* (i.e. *Cynipoidea* sp. and *Eurytoma melanagromyza*; Jayappa et al. 2002), as well for detail modelling of its potential economic impact in Brazilian soybean fields.

### 3.5 References


Arnemann JA, Guedes JVC, Stacke RF, Melo, A, Perini CR, Curioletti, LE. 2014. *Helicoverpa armigera*: até no inverno. *Cultivar Grandes Culturas* 183: 26-28.

Ascunce MS, Yang CC, Oakey J, Calcaterra L, Wu WJ, Shih CJ, Goudet TJ, Ross KG, Shoemaker D. 2011. Global invasion history of the fire ant *Solenopsis invicta*. *Science* 331: 1066-1068.

Ben-Dov E. 2014. *Bacillus thuringiensis* subsp. *israelensis* and its dipteran-specific toxins. *Toxins (Basel)* 6: 1222–1243.

Benson G. 1999. Tandem repeats finder: a program to analyse DNA sequences. *Nucleic Acids Research* 27: 573-580.

Bolsa de Cereales. 2015. Panorama Agrícola semanal. Estimaciones Agrícola 2015. [cited 05 September 2015]. Available from: <http://www.bolsadecereales.com>.

BRASIL, 2012. Instruçao Normativa n.7, from 13 April 2012. Ministério da Agricultura, Pecuária e Abastecimento (MAPA). [cited 05 September 2015]. Available from: <http://sistemasweb.agricultura.gov.br/pages/IN_MAPA_07_2012.pdf>.

Breniere SF, Bosseno MF, Vargas F, Yaksic N, Noireau F, Noel S, Dujardin JP, Tibayrenc M. 1998. Smallness of the panmictic unit of *Triatoma infestans* (Hemiptera: Reduviidae). *Journal of Medical Entomology* 35: 911-917.

Brier, H, Moore N. 2013. Soybean stem fly outbreak in soybean crops. The Beatsheet. [cited 05 September 2015]. Available from: <http://thebeatsheet.com.au/crops/pulses/soybeans/soybean-stem-fly-outbreak-in-soybean-crops/>.

Bueno AF, Sosa-Goméz DR. 2014. The old world bollworm in the Neotropical region: the experience of Brazilian growers with *Helicoverpa armigera*. *Outlooks on Pest Management*: 25: 1-4.

Campbell FT. 2001. The science of risk assessment for phytosanitary regulation and the impact of changing trade regulations. *BioScience* 51: 148-153.

Capinera JL. 2008. Encyclopedia of Entomology. 2nd Edition. Heidelberg, Springer, 4346pp.

Chung G, Singh RJ. 2008. Broadening the genetic base of soybean: a multidisciplinary approach. *Critical Reviews in Plant Sciences* 27: 295-341.

CONAB. Acompanhamento da Safra Brasileira – Grãos: oitavo levantamento, safra 2014/15. Brasília, p. 1-118, maio 2015.

Costa Lima AM, 1936. Dois curculionıideos daninhos no Rio Grande do Sul. *O Campo* 7: 23-24.

Czepak C, Albernaz KC, Vivan LM, Guimarães HO, Carvalhais T. 2013. Primeiro registro de ocorrência de *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae) no Brasil. *Pesquisa Agropecuária Tropical* 43: 110-113.

Davies N, Villablanca FX, Roderick GK. 1999. Bioinvasions of the medfly *Ceratitis capitata*: Source estimation using DNA sequences at multiple intron loci. *Genetics* 153: 351-360.

De Barro PJ, Liu SS, Boykin LM, Dinsdale AB. 2011. *Bemisia tabaci*: a statement of species status. *Annual Review of Entomology* 56: 1-19. doi: 10.1146/annurev-ento-112408-085504.

Dempewolf M. 2004. Arthropods of economic importance: Agromyzidae of the World. ETI- Information Services; Wokingham; UK (unpaginated).

Du YZ, Tang XT, Wang LP, Shen Y, Chang YW. 2014. Genetic differentiation of geographical populations of *Liriomyza sativae* (Diptera: Agromyzidae) in China based on mitochondrial COI gene sequences. *Mitochondrial DNA*: Early Online: 1-5. DOI: 10.3109/19401736.2014.987271.

FAOSTAT 2015. Statistical database. [cited 05 September 2015]. Available from: <http://faostat.fao.org/>.

Fletcher A. 2003. *Ophiomyia lantanae.* Research Station, Department of Natural Resources and Mines. The State of Queensland. [cited 05 September 2015]. Available from: <http://www.weeds.org.au/WoNS/lantana/docs/45_Ophiomyia_lantanae.pdf>.

Flux JEC. 1994. World distribution. In: The European Rabbit. The History and Biology of a Successful Colonizer (Eds. Thompson HV, King CM), pp. 8-21. Oxford University Press, Oxford.

Follett PA, Neven LG. 2006. Current trends in quarantine entomology. *Annual review of Entomology* 51: 359-385.

Gassen DK, Schneider S. 1985. Ocorrência de *Melanagromyza* sp. (Dip. Agromizidae) danificando soja no sul do Brasil. *In*: Reunião de Pesquisa da Soja da Região Sul, 13, 1985, Porto Alegre. Soja: resultados de pesquisa 1984-85. Passo Fundo: EMBRAPA-CNPT 108-109.

Giordano R, Cortez JCP, Paulk S, Stevens L. 2005. Genetic diversity of *Triatoma infestans* (Hemiptera: Reduviidae) in Chuquisaca, Bolivia based on the mitochondrial cytochrome b gene. *Mem Inst Oswaldo Cruz* 100: 753-760.

Guedes JVC, Arnemann JA, PERINI CR, Melo AA, Röhrig A. Stacke RF. Machado MRR. 2014. *Helicoverpa armigera*: da invasão ao manejo na soja. *Revista Plantio Direto* 137: 24-35.

Hardwick DF. 1965. The corn earworm complex. Ottawa: *Entomological Society of Canada* 5-247.

He L, Watabe H, Xiangyu J, Gao J, Liang X, Aotsuka T, Zhang Y. 2007. Genetic differentiation and cryptic speciation in natural populations of *Drosophila lacertosa. Molecular Phylogenetics and Evolution* 43: 24-31.

Hirose E, Moscardi F. 2012. Insetos de outras regiões do mundo. IN: EMBRAPA-Soja: manejo integrado de insetos e outros artrópodes-praga. Brasília, DF: Embrapa, 859 p.

Hunter ME. 2012. Using Genetic Research to Inform Imperiled and Invasive Species Management. U.S. Geological Survey Fact Sheet 3110. [cited 05 September 2015]. Available from: < http://pubs.usgs.gov/fs/2012/3017/pdf/FS_2012-3017_Hunter-WEB.pdf>.

IBGE 2015. Instituto Brasileiro de Geografia e Estatitica. Noticias. [cited 05 September 2015]. Available from: <http://www.ibge.gov.br/home/geociencias/geografia/default.shtm>.

Ivkosic SA, Gorman J, Lemic D, Mikac KM. 2014. Genetic monitoring of western corn rootworm (Coleoptera: Chrysomelidae) populations on a microgeographic scale. *Environmental Entomology* 43: 804-818. DOI: 10.1603/EN13264.

Jayappa AH, Reddy KMS, Kumar NG. 2002. Parasitoids of soyabean stem fly, *Melanagromyza sojae* (Zehntner) (Diptera:Agromyzidae). *Insect Environment* 8: 192.

Jeschke JM, Aparicia LG, Haider S, Heger T, Lortie CJ, Pysek P, Strayer DL. 2012. Support for major hypotheses in invasion biology is uneven and declining. *NeoBiota*, 14:1, pp. 1-20.

Kamiji T, Iwaizumi R. 2013. Four species of Agromyzidae (Diptera) intercepted by Japanese import plant quarantine. *Research Bulletin of the Plant Protection Service Japan* 49: 53-61.

Keane RM and Crawley MJ. 2002. Exotic plant invasions and the enemy release hypothesis. *Trends in Ecology & Evolution* 17:4, 164-170.

Kriticos DJ, Ota N, Hutchison WD, Beddow J, Walsh T, Tay WT, Borchert DM, Paula-Moreas SV, Czepak C, Zalucki MP. 2015. The potential distribution of invading *Helicoverpa armigera* in North America: is it just a matter of time? *PLoS ONE* 10(7): e0133224. doi:10.1371/journal.pone.0133224.

Leão TCC, Almeida WR, Dechoum MS, Leão T, Ziller SR, Almeida WR. 2011. Espécies Exóticas Invasoras no Nordeste do Brasil: Contextualização, Manejo e Políticas Públicas. CEPAN-Instituto Hórus. Recife, 99p.

Leite NA, Alves-Pereira A, Correa AS, Zucchi MI, Omoto C. 2014. Demographics and genetic variability of the New World bollworm (*Helicoverpa zea*) and the Old World bollworm (*Helicoverpa armigera*) in Brazil. *PLoS One* 9: 11 e113286.

Librado P, Rozas J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452.

Link D, Hatschbach M, Link FM. 2008. Mosca da haste da soja: ocorrência e danos. In: XXXVII Reunião de Pesquisa de Soja da Região Sul, Porto Alegre-RS.

Lopes-da-Silva M, Sanches MM, Stancioli AR, Alves G, Sugayama R. 2014. The role of natural and human-mediated pathways for invasive agricultural pests: a historical analysis of cases from Brazil. *Agricultural Sciences* 5: 634-646.

Lowe SJ, Browne M, Boudjelas S. 2001. 100 of the world's worst invasive alien species. *Aliens* 12: 1-14.

Lukefahr MJ, Barbosa S, Braga RS. 1994. The introduction and spread of the boll weevil, *Anthonomus grandis* Boh. (Coleoptera) in Brazil. *Southwestern Entomologist* 19: 414-418.

Macdonald GK, Brauman KA, Sun S, Carlson KM, Cassidy ES, Gerber JS, West PC. 2015. Rethinking agricultural trade relationships in an era of globalization. *BioScience* 65: 275-289. doi: 10.1093/biosci/biu225.

Mack RN, Chair DSW, Lonsdale M, Evans H, Clout M, Bazzaz F. 2000. Biotic invasions: causes, epidemiology, global consequences and control. *Ecological Application* 10: 1-20.

Mastrangelo T, Paulo D, Bergamo L, Morais E, Silva M, Bezerra-Silva G, A.M.L. Azeredo-Espin. 2014. Detection and genetic diversity of a Heliothine invader (Lepidoptera:

Noctuidae) from north and northeast of Brazil. *Journal Economic Entomology* 107: 970-980.

Mata D, Freitas RE. 2008. Produtos agropecuarios: para quem exportar? *Revista de Economia e Sociologia Rural* 46: 257-290.

McKercher B, McGowan JA, Howarth JK, Saito A. 1953. Preliminary report on the isolation and identification of bluetongue virus from sheep in California. *Journal of the American Veterinary Medical Association* 122: 300-301.

Michereff MF, Vilela EF. 2001. Traça-do-tomateiro, *Tuta absoluta* (Lepidoptera: Gelechiidae). In: Historico e impacto de pragas introduzidas no Brasil. Ed. by Vilela E, Zucchi RA, Cantor F, Holos, Ribeirão Preto, Brazil 81-84.

Monteiro RC, Zucchi RA, Mound LA, 1995. Record of *Thrips palmi* (Thysanoptera, Thripidae) in the State of Sao Paulo, Brazil. *Revista Agricola* 70:53–55.

Navia D, Marsaro Jr AL. 2010. First report of the citrus Hindu mite, *Schizotetranychus hindustanicus* (Hirst) (Prostigmata: Tetranychidae), in Brazil. *Neotropical Entomology* 39: 140-143.

NCBI 2015. National Center for Biotechnology Information, U.S. National Library of Medicine. [cited 05 September 2015]. Available from: <http://www.ncbi.nlm.nih.gov>.

Nimer E. 1979. Um modelo metodologico da classificaccao de climas. *Revista Brasileira de Geografia* 4: 59-89.

NPAG. 2000. *Melanagromyza obtusa:* Pigeonpea pod fly. Potential threat to pigeonpea in Puerto Rico. [cited 05 September 2015]. Available from: <http://www.pestalert.org/storage/dipagrmo300.pdf>.

Oliveira CM, Auad AM, Mendes SM, Frizzas MR. 2012. Economic impact of exotic insect pests in Brazilian agriculture. *Journal Applied Entomology* 137:1–15.

Pailey PT. 2007. Pests of field crops and pastures: identification and control. CSIRO Publishing, Collingwood VIC, Australia.

Pomari-Fernandes A, Bueno AF, Sosa-Gómez D. 2015. *Helicoverpa armigera*: current status and future perspectives in Brazil. *Current Agricultural Science and Technology* 21: 1.

Purse BV, Mellor PS, Rogers DJ, Samuel AR, Mertens PPC, Baylis M. 2006. Climate change and the recent emergence of bluetongue in Europe. *Nature Reviews Microbiology* 4: 160.

Ramalho FS, Santos RF. 1994. Impact of the introduction of the cotton boll weevil in Brazil. In: Proceedings of the World Cotton Research Conference-1: Challenging the future. Ed. by Constable GA, Forrester NW, CSIRO, Brisbane, Australia, 466-474.

Reaser JK, Galindo-Leal C, Ziller SR. 2005. Visitas indesejadas: a invasão de espécies exóticas. In: Galindo-Leal C.; Câmara IDG. (Eds.). Mata Atlântica: biodiversidade, ameaças e perspectivas. Fundação S.O.S. Mata Atlântica: São Paulo. Conservação Internacional: Belo Horizonte registro de ocorrência de *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae) no Brasil. *Pesquisa Agropecuária Tropical* 43: 110-113.

Roque A, Auger-Rozenberg MA. 2006. Tentative analysis of the interceptions of non-indigenous organisms in Europe during 1995 – 2004. *Bulletin OEPP/EPPO Bulletin* 36: 490-49.

Scheffer SJ, Winkler IS, Wiegmann BM. 2007. Phylogenetic relationships within the leaf-mining flies (Diptera: Agromyzidae) inferred from sequence data from multiple genes. *Molecular Phylogenetic and Evolution* 42: 756-75.

Scheffer SJ. 2000. Molecular evidence of cryptic species within the *Liriomyza huidobrensis* (Diptera: Agromyzidae). *Journal of Economic Entomology* 93: 1146-1151.

Shepard BM, Barrion AT. 1998. Parasitoids of insects associated with soybean and vegetable crops in Indonesia. *Journal of Agricultural Entomology* 15: 239-272.

Shepard, M. Lawn RJ. Schneider MA. 1983. Insects on Grain Legumes in Northern Australia: A Survey of Potential Pests and Their Enemies. University of Queensland Press. Australia. 81pp.

Short J, Smith AP. 1994. Mammal decline and recovery in Australia. *Journal of Mammalogy* 75: 288-297.

Shrivastava AN, Singh CB, Shrivastava MK. 2002. Soybean. In C. B. Singh and D. Khare (Eds.), Genetic improvement of field crops. (pp. 136-157). Jodhpur, India: Scientific Publishers.

Silva AGA, Gonçalves CR, Galvão DM, Gonçalves AJL, Gomes J, Silva MN, Simoni L. 1968. Quarto catálogo dos insetos que vivem nas plantas do Brasil. Seus parasitos e predadores. Parte 2, Tomo 1, insetos, hospedeiros e inimigos naturais. Rio de Janeiro, Ministério da Agricultura, 622pp.

Spencer KA. 1973. Agromyzidae (Diptera) of Economic Importance. Series Entomologica. Dr. W. Junk B. V., Hague the Netherlands. v.9, 418p. 1973.

Staden RBKF, Bonfield JK. 2000. The Staden package, 1998. *Methods in Molecular Biology* 132: 115-130, PMID10547834. doi:10.1385/1-59259-192-2:115.

Strakhova IS, Yefremova ZA, Tschirnhaus M, Yegorenkova EN. 2013. The parasitoid complex (Hymenoptera, Eulophidae) of leafminer flies (Diptera, Agromyzidae) in the middle Volga Basin. *Entomological Review* 93(7): 865-873. DOI: 10.1134/S0013873813070087.

Talekar NS. 1989. Characteristics of *Melanagromyza sojae* (Diptera: Agromyzidae) damage in soybean. *Journal Economic Entomology* 82: 584-588.

Tamura K, Nei M, Kumar S. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences* 101: 11030-11035.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* 30: 2725-2729.

Tay WT, Beckett SJ, De Barro PJ. 2014. Phosphine resistance in Australian *Cryptolestes* species (Coleoptera: Laemophloeidae): Perspective from mitochondrial DNA Cytochorme Oxidase I analysis. *Pest Management Science*. doi: 10.1002/ps.3805. [Epub ahead of print].

Tay WT, Soria MF, Walsh T, Thomazoni D, Silvie P, Behere GT, Anderson C, Downes S. 2013. A brave new world for an Old World pest: *Helicoverpa armigera* (Lepidoptera: Noctuidae) in Brazil. *PLoS ONE* 8(11): e80134. doi:10.1371/journal.pone.0080134

Tay WT, Evans GA, Boykin LM, De Barro PJ. 2012. Will the real *Bemisia tabaci* please stand up? *PLoS One* 7(11):e50550.

Templeton AR, Crandall KA, Sing CF. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132: 619-633.

Thapa RB. 2012. Redescription of *Melanagromyza sojae* (Zehntner) from India and Nepal. *Nepalese Journal of Biosciences* 2: 64-70.

USDA (United States Department of Agriculture). List of intercepted plant pests, 1940. Bureau of Entomology and Plant Quarantine, Florida, Issued December 1941.

Van Beurden EK. 1981. Bioclimatic limits to the spread of *Bufo marinus* in Australia: a baseline. *The Proceedings of the Ecological Society of Australia* 11: 143-149.

Van Den Berg H, Ankasah D, Hassan K, Muhammad A, Widayanto Ha, Wirasto Hb, Yully I. 1995. Soybean stem fly, *Melanagromyza sojae* (Diptera: Agromyzidae), on Sumatra: Seasonal incidence and the role of parasitism. *International Journal of Pest Management* 41:3 127-133. DOI:10.1080/09670879509371937

Verma SN, Singh OP, Thakur RC. 1989. A new record of soybean stem borer as a pest of urdbean in Madhya Pradesh. *Indian Journal of Pulses Research* 2: 181-183.

Vidal O, García-Berthou E, Tedesco PA, García-Marín JL. 2010. Origin and genetic diversity of mosquitofish (*Gambusia holbrooki*) introduced to Europe. *Biological Invasions* 12: 841-851.

Wang CL. 1979. Occurrence and life-history of *Melanagromyza sojae* on soybean. *Journal of Agricultural Research of China* 28: 217-223.

Wang J, Gai J. 2001. Mixed inheritance model for resistance to agromyzid beanfly (*Melanagromyza sojae* Zehntner) in soybean. *Euphytica* 122: 9-18.

Winkler IS, Scheffer SJ, Mitter C. 2009. Molecular phylogeny and systematics of leaf-mining flies (Diptera: Agromyzidae): delimitation of *Phytomyza* Fallén sensu lato

and included species groups, with new insights on morphological and host-use evolution. *Systematic Entomology* 34: 260–292. doi:10.1111/j.1365-3113.2008.00462.x

Yorinori JT,Paiva WM, Frederick RD, Costamilan LM, Bertagnolli PF, Hartman GE, Godoy CV, Nunes Junior J. 2005. Epidemics of soybean rust (Phakopsora pachyrhizi) in Brazil and Paraguay from 2001 to 2003. *Plant Disease* 89: 675-677.

Yu H, Romeis J, Li Y, Li X, Wu K. 2014. Acquisition of Cry1Ac protein by non-target arthropods in Bt soybean fields. *PLoS ONE* 9(8): e103973. doi:10.1371/journal.pone.0103973

Zhan S, Zhang W, Niitepõld K, Hsu J, Fernández Haeger J, Zalucki MP, Altizer S, Roode JC, Reppert SM, Kronforst MR. 2014. The genetics of monarch butterfly migration and warning colouration. *Nature* 514: 317-321.

Zhang DX, Hewitt GM. 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology* 12: 563-584.

Ziaee M. 2012. Oilseed Pests. InTech. [cited 05 September 2015]. Available from: <http://www.intechopen.com/books/oilseeds/oilseed-pests>.

# CHAPTER IV

## MOLECULAR DETECTION AND GENETIC DIVERSITY OF THE OLD WORLD COTTON BOLLWORM *HELICOVERPA ARMIGERA* (LEPIDOPTERA: NOCTUIDAE) IN CONE SUL OF AMERICA

## 4.1 Introduction

The risk of exotic species entering Brazil is high as Brazilian agriculture is characterized by its vast landmass, the country's extensive border (15.7 thousand kilometres) neighbouring 10 countries, and its large export and import flows. These factors have created numerous opportunities as entry points for exotic organisms (Oliveira et al. 2012; IBGE 2015). Insects in particular, are a risk because of their small size and ability to survive unfavourable conditions during transport and storage of agricultural produce and plant propagation materials. In addition, insects often have a high capacity for dispersal, high reproduction rates, short generation time, and their ability to colonize new environments after arrival (Kiritani and Yamamura 2003; Bounfour et al. 2005).

The Old World cotton bollworm *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae) is a highly polyphagous pest of several economically significant crops, and is one of the most widely distributed pest species of the Heliothinae group (Fig. 1) (Tay et al. 2013; King, 1994). *H. armigera* is widely distributed in almost all continents except North America and Antarctica though recently a small number of individuals have been detected in Puerto Rico and in Florida (USDA/APHIS 2015).

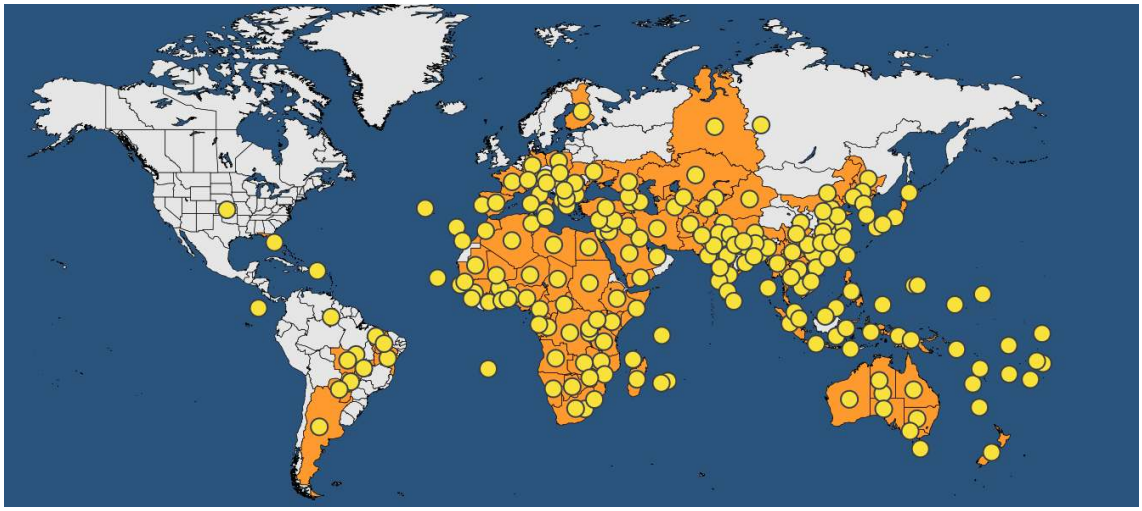

**Fig. 1:** The current global distribution of *H. armigera*. (Source: adapted from EPPO 2015, 23 Sep 2015).

The confirmation of the presence and spread of *H. armigera* in South and Central America, e.g., in Brazil in 2013 (Czepak et al. 2013; Tay et al. 2013), in Paraguay in 2013 (SENAVE 2013), and in Argentina in 2014 (Murúa et al. 2014), has serious implications in terms of the management of this pest. Kriticos et al. (2015) recently modelled the extraordinary spread potentials of this pest into North America assuming a South American source population, and within four months post Kriticos et al. (2015) paper publication, USDA/APHIS and Florida Department of Agriculture and Consumer Services (FDACS) confirmed the detection of *H. armigera* in mainland USA (Florida, July-2015, USDA), with confirmation of its detection in the US offshore territory of Puerto Rico only nine months earlier (9-October-2014) (USDA).

The recent detection of *H. armigera* in Brazil and the United States of America demonstrates the speed at which an invasive pest can become established and spread through the landscape. *H. armigera* was first identified in Brazil during the 2012/13 crop season (Czepak et al. 2013; Specht et al. 2013; Tay et al. 2013). This new pest triggered a national phytosanitary emergency in Brazil, where it attacked a number of

different crops (e.g., cotton, soybean, corn, green beans, tomatoes, citrus, etc.) and pastures (Bueno and Sosa-Goméz 2014), that resulted in significant economic costs (estimated at between USD$0.8 billion to USD$2 billion damages in the (2012/13) cropping season in Brazil; Guedes et al. 2014; Bueno and Sosa-Goméz 2014; Mastrangelo et al. 2014; Lopes-da-Silva et al. 2014). Intriguingly, *H. armigera* was reported to be widespread in the country in 2012/13, a distribution pattern that might otherwise be expected to take many months or even years to occur, damaging summer and winter crops (Bueno and Sosa-Goméz 2014; Arnemann et al. 2014). Subsequent investigations into population-wide genetic diversity based on both multiple nuclear and mitochondrial DNA (mtDNA) loci (Mastrangelo et al. 2014) and the mtDNA Cytochrome Oxidase subunit I (COI) locus (Leite et al. 2014) identified unexpectedly high population diversity in this new invasive pest, a pattern contrary to the 'classical' biosecurity incursion genetic diversity pattern that involved a single founder (or low number of founders) effect (e.g., *Diabrotica virgifera virgifera* in Europe, Ivkosic et al. 2014). Various factors underlying and/or contributing to this invasion and rapid build-up of *H. armigera* populations in Brazil have been proposed that included multiple incursions (e.g., Tay et al. 2013; Leite et al. 2014), and the long distance dispersal ability of *H. armigera* (e.g., Leite et al. 2014), although association with human activities (e.g., international trade) or through chance natural incursion were considered plausible but unlikely (Lopes-da-Silva et al. 2014).

Although first reported in February 2013, the widespread presence of *H. armigera* led to the belief that this pest was already present in Brazil but remained undetected for some time. This was in part, due to its morphological similarity to other Heliothinae species (Pomari-Fernandes et al., 2015) such as the closely related *H. zea*

(Hardwick 1965). Behere et al. (2007), Behere (2008), and Tay et al. (2013) surveyed a population of 30 *H. zea* collected from Mato Grosso (MT) state in 2006 using mitochondrial markers (i.e., mtCOI and/or Cytochrome *b* (Cyt*b*)) and failed to detect the presence of *H. armigera*, potentially suggesting either the absence of *H. armigera* in MT (and in Brazil) and/or possibly due to the combined effects of the small number of specimens used by Behere et al. (2007), and a much smaller *H. armigera* population size at the time (as would be expected during the introduction, and early lag-phases of the pest) (i.e., stochastic lineage sorting during the lag-phase of exotic species incursion, Mack and Erneberg (2002); Schoener and Schoener (1983)).

Alternatively, the introduction of *H. armigera* could have taken place in the period after the collections as reported in Behere et al. (2007) and Behere (2008). Sequencing of the mitochondrial DNA cytochrome oxidase (mtDNA COI) gene from a small number of suspected 2013 field-collected individuals from MT subsequently confirmed its presence (Tay et al. 2013). A well-established and characterized database of the mtDNA genes such as the COI and Cyt*b* genes (e.g., De Barro et al. 2011 for *Bemisia tabaci*; Behere et al. 2007; Tay et al. 2013 for *H. armigera*) therefore underpinned the efficiency for molecular identification of invasive exotic species, as exemplified by the confirmation of *H. armigera* in the South America.

Accurate and rapid molecular identification of pest species will further enable appropriate management measures to be developed and be implemented in a timely manner, such as developing strategies to minimize the spread of insect pests across agricultural landscapes, and maximizing pest control options such as introducing appropriate biological control agents including natural enemies (e.g., parasitoids). Hebert et al. (2004) proposed using of the N-terminal (5') region (648pb) of the mtDNA

COI gene as the core "DNA barcode" of a global identification system for metazoans and showed a model mtDNA COI profile that was 100% successful in correctly identifying 200 specimens from closely allied species of lepidopterans. Others authors have also reported success using the C-terminal region (i.e., 5' end) of the COI and other (e.g., Cyt*b*) genes to study population structure in Lepidoptera (Behere et al. 2007; Behere 2008; Albernaz et al. 2012; Tay et al. 2013). Molecular diagnostic tests are widely accepted as an essential component of detection and identification systems of exotic invasive species, e.g., the Australian soybean moth (*Aproaerema simplixella*) in Africa, Buthelezi et al. (2012), the soybean stem fly *Melanagromyza soja* (Arnemann et al. 2015; Chapters II and III).

Information about the geographical distribution, invasion sources and dispersal routes of *H. armigera* remained unknown in Brazil. Because first reports confirming the presence of *H. armigera* in the South American continent was in Brazil (Czepak et al. 2013, Tay et al. 2013), and owing to the common perception that *H. armigera* has the capability to accomplish long distance migration especially under favourable weather conditions, and also that neighbouring countries (e.g., Argentina, Paraguay) have only reported the appearances of *H. armigera* within their agricultural landscapes (SENAVE 2013; Murúa et al. 2014) after morphological and molecular confirmation in Brazil, might suggest *H. armigera* incursions in these southern regions of South America (i.e., Cone Sul) potentially had their source of populations originated from within Brazil.

Here, I analysed the genetic sequence data from the mtDNA partial gene of *H. armigera* specimens collected from Uruguay, Argentina, Paraguay, and also from three southern Brazilian states (Paraná; Rio Grande do Sul and Santa Catarina) with the aim to better characterise the genetic diversity of *H. armigera* in the Cone Sul regions of

Latin America. Results from this study confirmed, at the molecular level, that *H. armigera* is present in soybean fields in Uruguay and Paraguay, and also in Rio Grande do Sul (RS), Santa Catarina (SC) and Paraná (PR) states. This study, representing also the first molecular genetic diversity characterisation of *H. armigera* in the Cone Sul region (and including individuals from Argentina), also detected unexpected unique haplotype diversity in *H. armigera* populations from non-Brazil sites. Analyses of the South American spatial mtDNA COI diversity patterns were also carried to infer possible introduction scenarios that might explain the observed spatial heterogeneous mtDNA COI data, and to enable discussion on potential biosecurity threats to the Cone Sul territory by this pervasive polyphagous pest.

## 4.2 Materials and methods

### 4.2.1 Samples

Adults of suspected *H. armigera* were collected using delta traps baited with the female sexual pheromone Iscalure armigera® (ISCA Tecnologias LTDA, Ijuí, RS, Brazil) randomly installed in soybean fields in Rio Grande do Sul (RS), Santa Catarina (SC) and Paraná (PR) states from Brazil, and also in Uruguay, Argentina and Paraguay, in the cropping 2014/15 season (Table 1). Adult moths were collect from the traps and preserved in individual 1.5mL Eppendorf tubes with 1.0 mL of 99.9% ethanol for molecular characterization purposes.

**Table 1:** Collection sites, dates, and mtDNA COI GenBank accession numbers of *Helicoverpa armigera* specimens from Brazil, Uruguay (URU), Argentina (ARG) and Paraguay (PRY). Brazilian States are Rio Grande do Sul (RS), Santa Catarina (SC) and Paraná (PR).

| Sample ID | Code/state | Geographical coordinates Lat/Lon | Sampling date | GenBank Accession No. |
|---|---|---|---|---|
| 1-4 | ARG | -32.1963222; -61.716597222 | 09-April-2014 | |
| 5-14 | PRY | -25.25508055; -57.56711111 | 12-April-2014 | |
| 15-16 | URY | -34.904344444; -54.93699166 | 03-May-2014 | |
| 17-28 | BRA-RS | -29.72653611; -53.561161111 | 10-April-2014 | |
| 29-34 | BRA-SC | -26.462400; -53.5098694444 | 06-April-2014 | |
| 35-37 | BRA-PR | -24.1583361111; -49.821575 | 25-March-2014 | |

### 4.2.2 Total genomic DNA (gDNA) extraction

The total gDNA from either 2-3 legs and/or partial abdomens of adult moths was extracted at the CSIRO Black Mountain Laboratories (ACT, Australia) using Qiagen Blood and Tissue DNA extraction kit (Cat# 69506) and following the protocol as provided by the manufacturer. The extracted gDNA was eluted in 50µL of buffer AE and stored at -20˚C until needed.

### 4.2.3 Primers, PCR amplification and sequencing of partial mitochondrial DNA COI gene

A 707 bp fragment of the mtDNA COI gene was initially PCR amplified using the primers NOC-COI-F (5' GCGAAAATGACTTTATTCAAC 3') and COI-R (5' GCGAAAATGACTTTATTCAAC 3') (unpublished data, W.T. Tay CSIRO). For PCR the following conditions were used: denaturing at 95°C for 5 minutes, 34 cycles of alternating denaturing, annealing and extension steps each for 30 seconds at 95°C, 61°C and 72°C respectively, followed by a final extension cycle of 72°C for 5 minutes.

PCR amplification of individual DNA samples was carried out in a total reaction volume of 25 µL, and contained 25 ng of genomic DNA, 0.5 µM of both forward and reverse primers, 0.2 mM of dNTP's, 1× Phusion HF Buffer (NEB), and 1.25 units of Phusion DNA polymerase (NEB).

Amplicons were purified using the QIAquick® PCR purification Kit (Qiagen) prior to being used as DNA template for Sanger sequencing reaction using the ABI BigDye® dideoxy chain termination sequencing system V3.1 (Applied Biosystems). Sequencing reaction and post sequencing reaction clean-up was as required by the sequencing facility. Sequencing of all partial mtDNA COI gene region was carried out at the Australian National University Biomolecular Resource Facility (ANU BRF).

### 4.2.4 Sequence analysis of partial mtDNA COI gene

The programs Pregap and Gap4 within the Staden package (Staden et al. 2000) were used for editing and analysing the DNA sequences and to generate sequence contigs. Assembled partial mtDNA COI contigs were checked for premature stop codons that may indicate pseudogenes using Geneious® R8 (Biomatters Ltd., New Zealand) and Blastp (for amino acid homology confirmation) to available *H. armigera* partial mtDNA COI genes in public database (e.g., GenBank, iBoL). Nucleotide sequences without premature stop codons and that differed by one or more nucleotides were considered as different haplotypes, while sequences exhibiting identical SNPs at the same nucleotide positions were considered as the same haplotype.

### 4.2.5 MtDNA COI haplotype network and evolutionary divergence

A mtDNA COI haplotype network for *H. armigera* was constructed using the statistical parsimony phylogenetic network estimation program TCS v1.21 (Templeton et al. 1922) and redrawn by hand. The distribution map of the *H. armigera* haplotypes was build using PopArt <http://popart.otago.ac.nz>. For analysis the complete dataset was partitioned into 4 groups: (i) Asia (India, Pakistan and China), (ii) Europe (Germany and unknown sites in Europe), (iii) Australia, and (iv) South America (Brazil, Argentina, Paraguay and Uruguay) (Table 2). Levels of polymorphism between countries/continents were estimated by calculating both haplotype diversity (*h*, the probability that two randomly chosen haplotypes are different in the sample) and nucleotide diversity ($\pi$, the average number of nucleotide differences per site between two sequences) (Nei 1987). Both *h* and $\pi$ were estimated using the DnaSP software by Librado and Rozas (2009) (see Table 3).

Categorization of global mtDNA COI haplotypes at the 5' gene region (see Table 4), and estimates of evolutionary divergence between all *H. armigera* individuals (i.e., between South America *vs.* Europe *vs.* Asia *vs.* Australia; n=314) involved 548bp of the mtDNA COI partial gene in the final dataset. The evolutionary divergence and the average intra-species groups' evolutionary diversity were estimated using the Maximum Composite Likelihood model (Tamura et al. 2004), and included all (i.e., $1^{st}$+$2^{nd}$+$3^{rd}$) codon positions, while gaps and missing data were excluded. Evolutionary genetic analyses were conducted in MEGA6 (Tamura et al. 2013).

**Table 2:** South American and global *H. armigera* sequences and iBoL/GenBank accession numbers for the mtDNA COI 5' region. Numbers of individuals (n) from each location are indicated in parentheses.

| Countries or Continents | Locations (n) | iBoL/GenBank accession numbers |
|---|---|---|
| **China** | Kunming (8) | GQ892840.1, GQ892842.1, GQ892854.1, GQ995232.1, GQ995234.1, GQ995235.1, GQ995244.1, GQ995239.1 |
| | Tibet (2) | JX392497.1, JX392415.1 |
| | Miaofengshan (4) | JX509766.1, JX509765.1, JX509764.1, JX509739.1 |
| | Not specified (1) | HQ132369.1 |
| | Dali (15) | GQ892846.1, GQ892847.1, GQ892848.1, GQ892849.1, GQ892850.1, GQ892851.1, GQ892852.1, GQ892853.1, GQ995233.1, GQ995236.1, GQ995237.1, GQ995240.1, GQ995241.1, GQ995242.1, GQ995243.1 |
| | Henan (2) | GQ995238.1, GQ892855.1 |
| | Lijiang (1) | GQ892845.1 |
| | Yuxi (2) | GQ892843.1, GQ892844.1 |
| **Thailand** | Not specified (1) | EU768935.1 |
| **India** | Not specified (6) | HM854928.1, HM854929.1, HM854930.1, HM854931.1, HM854932.1, JX532104.1 |
| **Pakistan** | Not specified (2) | JN988529.1, JN988530.1 |
| **Europe** | Not specified (24) | FN907979.1, FN907980.1, FN907988.1, FN907989.1, FN907995.1, FN907996.1, FN907997.1, FN907998.1, FN907999.1, FN908000.1, FN908001.1, FN908002.1, FN908003.1, FN908005.1, FN908006.1, FN908011.1, FN908013.1, FN908014.1, FN908015.1, FN908016.1, FN908017.1, FN908018.1, FN908023.1, FN908026.1 |
| | Germany (4) | GU654969.1, GU686757.1, GU686955.1, JF415782.1 |
| **Australia** | Queensland (10) | ANICL283-10, ANICL284-10, GBGL12943-14, GBGL12944-14, GBGL12945-14, GBGL12946-14, GBGL12947-14, GBGL12948-14, GBGL12949-14, GBGL12950-14 |
| | New South Wales (8) | GBGL12935-14, GBGL12936-14, GBGL12937-14, GBGL12938-14, GBGL12939-14, GBGL12940-14, GBGL12941-14, GBGL12942-14 |
| **Brazil** | Bahia (112) | KM274936-KM274938, KM27513-KM275140 KM274939-KM274941, KM274943-KM274950, KM274951-KM274953, KM274957- |

| | | |
|---|---|---|
| | | KM274975, KM274979-KM274986, KM275038-KM275052, KM275078-KM275082, KM275070-KM275077, KM275127-KM275136, KF624850-KF624861 |
| | Maranhão (20) | KM27498-KM274996, KM275103-KM275112 |
| | Mato Grosso (19) | KM275083-KM275092, KM275156-KM275158, KM275097-KM275102 |
| | Piaui (39) | KF624811- KF624849 |
| | Roraima (14) | KF624862- KF624875 |
| | Parana (2) | This study |
| | Santa Catarina (7) | This study |
| | Rio Grande do Sul (12) | This study |
| **Argentina** | Santa Fé (4) | This study |
| **Uruguay** | Maldonado (2) | This study |
| **Paraguay** | Alto Paraná (10) | This study |

### 4.2.6 Analysis of *Helicoverpa armigera* haplotypes distributions

To better investigate the spatial distribution patterns of *H. armigera* haplotypes in the South American continent, a matrix table was prepared for the frequencies of the 25 *H. armigera* mtDNA COI haplotypes identified to-date from 11 South American locations (i.e., from Brazilian samples sites: Bahia (BA), Maranhão (MA), Mato Grosso (MT), Piaui (PI), Roraima (RR), Paraná (PR), Santa Catarina (SC) and Rio Grande do Sul (RS); and from Argentina, Uruguay, and Paraguay) (Table 5), prior to performing a contingency table analysis using the $\chi^2$ statistic to detect departures, as detailed below. A test matrix that involved randomisation of haplotypes x locations was also generated according to an appropriate null model (see 'null model' below), for exploring the probability of departures of observations from the null model (e.g., similar to Gotelli's (2000) analysis of species co-occurrence data).

### 4.2.7 Null and alternative hypotheses

Due to the perceived unevenness of mtDNA COI haplotype spatial patterns, it would be ideal if one could ascertain whether the diversity and frequencies of individual haplotypes had occurred independently and at random across South American sampling sites. For this purpose, the null hypothesis is therefore that mtDNA COI haplotypes are being independently sampled and are distributed randomly across the landscapes. The alternative hypothesis therefore considers at least one haplotype as being either more or less common, in at least one location, than expected due to chance alone (i.e. haplotypes were differentially distributed across the landscapes).

### 4.2.8 Test statistics used

Equation (1) below was used for the $\chi^2$ statistic to detect departures from randomness:

$$\chi^2_{Obs} = \sum_{i,j} \frac{\left(Z_{Obs_{i,j}} - \bar{Z}_{Ran_{i,j}}\right)^2}{\bar{Z}_{Ran_{i,j}}} \qquad (1)$$

where $Z_{Obs_{i,j}}$ is the observation for row *i* and column *j* within the matrix (i.e. Table 5), and $\bar{Z}_{Ran_{i,j}}$ is the expected value for the same table entry calculated under the null hypothesis. If certain haplotypes and/or locations were disproportionately under- or over- represented, then $\chi^2_{Obs}$ would be expected to fall within the extreme tail of the distribution of $\chi^2_{Ran}$ (i.e., the calculated value of the statistic when the null hypothesis is known to be true).

Although an analysis based on $\chi^2_{Obs}$ provides an overall test of haplotype randomness across the landscape, information on which location and/or haplotype

combinations underpin any deviation from randomness will require an additional test statistic (i.e., equation (2)) to be applied at the scale of each haplotype/location combination:

$$TS_{DIFF} = Z_{Obs_{i,j}} - \bar{Z}_{Ran_{i,j}} \quad (2)$$

Equation (2) implies that when $TS_{DIFF}$ < 0.0 a haplotype would be less common than expected in location *j* due to chance alone, when $TS_{DIFF}$ > 1.0 the haplotype would be more common than expected in location *j* by chance, and when $TS_{VAR} = 0.0$ the observed data conformed with the null model.

### 4.2.9 Null model

The null model generated for this study involved distributing all 226 observations (i.e., the complete mtDNA COI dataset of *H. armigera* in South America; Table 5) across haplotypes and locations at random, but also specifically under the additional constraint that the row and column totals must remain fixed (i.e. the same number of observations of each haplotype, and the same number of observations for each location, are both retained in the randomised matrices). This treatment is necessary to ensure that the null model explicitly accounts for both unequal survey efforts at different sites and the overall lower frequencies of some haplotypes. The algorithm AS159 of Patefield (1981) was able to generate random tables constraining both row and column marginal totals, and was therefore used to implement the null model (based on 10,000 randomly generated tables).

### 4.2.10 Calculating P-values

P-values for $\chi^2_{Obs}$ were calculated by enumerating the number of times $\chi^2_{Ran}$ was less than or equal to $\chi^2_{Obs}$ ($N_{LessEqual}$), and also the number of times $\chi^2_{Ran}$ was greater than or equal to $\chi^2_{Obs}$ ($N_{greaterEqual}$) based on 10,000 simulations. The P-value is then the smaller of these two quantities divided by 10,000, and multiplied by two to affect a two-tailed test (Manly 2001) (i.e., corresponds to a haplotype being able to either be more or less common in the landscapes) (equation 3):

$$P\text{-}value = \frac{\min(N_{GreaterEqual}, N_{LessEqual})}{10{,}000} \times 2 \quad (3)$$

P-values for $TS_{DIFF}$ were calculated similarly at the scale of individual table cells. Note that as the quantities $N_{GreaterEqual}$ and $N_{LessEqual}$ can potentially overlap in their included values (i.e., both include an equality term), the P-value can therefore potentially exceed 1.0, and in which case the values are rounded down to 1.0.

### 4.2.11 Validating the null model

The Null models was validated by creating *n* pseudo-observed data sets at random to confirm appropriate type I and type II error rates were obtained. To achieve this, *n* pseudo-observed data sets at random (i.e. data sets consistent with the null hypothesis) was created and applied to each of the test. At the 5% confidence level only 5% of pseudo random data sets should be statistically significant, and the expected P-value distribution across the *n* tests should be rectangular. Validation of the null model will enable elevated rates of type I error (i.e., a chance that when applying the test, significant differences could be concluded even when no real

differences are actually present). If it is found that more than 5% are significant this will indicate an elevated type I error rate (i.e., a chance that the null hypothesis is incorrectly rejected when no real differences are actually present). Conversely, when less than the nominal number of pseudo random data sets yield a significant result it can lead to incorrect acceptance of the null hypothesis (i.e., concluding no significant difference even when one is actually present; (i.e., an elevated Type II error rate)). Elevated Type II error rates are generally considered less serious than elevated type I error rates.

The null model was validated through repeated analyses (10,000 times) using random pseudo-data reconstructed to match the n = 266 observations in Table 5, but allowing observations to be randomly allocated to the matrices, and thereby ensuring conformity with the null hypothesis. Tables that had no observations for a given row or column during the constructing of pseudo-data were excluded.

### 4.2.12 Sub-table analysis

Analysis of $TS_{DIFF}$ can be used to further explore the primary responsible circumstances (i.e., combinations of haplotype and location) and directionality (i.e., whether haplotypes were unexpectedly rare or common across the sites) if/when evidence of differentially distributed haplotypes across the landscapes was detected (i.e. to identify the genotypes that are unduly rare or common across the locations).

### 4.2.13 Brazil vs. Non-Brazil

The above analyses were repeated, but combining the location data to consider just two categories – Brazil and Non-Brazil samples.

## 4.3 Results

### 4.3.1 PCR amplification and sequence analysis

All specimens from the southern/south-western regions of South America (e.g., Argentina, Uruguay, Paraguay, Paraná (Brazil), Rio Grande do Sul (Brazil) and Santa Catarina (Brazil)) were successfully sequenced for the mtDNA COI fragment using the Noc-COI-F/R primer pairs. Sequence identity search against the NCBI GenBank database confirmed that all suspected moths significantly matched (i.e., 99-100% nucleotide identity) published *H. armigera* sequences.

The range of genetic distances of *H. armigera* within Asia (China, India and Pakistan) and within Australia were both 0.00 – 0.04%, while within Europe (Germany and unknown sites) and within South America (Brazil, Argentina, Uruguay and Paraguay) were both 0.00 - 0.02%. Estimates of evolutionary divergence between all *H. armigera* sequences from Australia, Asia, Europe and South America were therefore likewise low and ranged from 0.00 - 0.04%. Observed nucleotide diversity between countries/continents in *H. armigera* ranged from 0.00238 ± 0.0004 (s.e.) – 0.00403 ± 0.00057 (s.e.) (Table 3).

**Table 3**: Comparison of *Helicoverpa armigera* partial mtDNA COI gene nucleotide diversity ($\pi$ ± s.e.) and haplotype diversity (h ± s.e.) between different countries/continents.

| Location | Nucleotide diversity ($\pi$) | Haplotype diversity (h) |
|---|---|---|

| | | |
|---|---|---|
| **Asia** | 0.00360 ± 0.00034 | 0.912 ± 0.024 |
| **Europe** | 0.00238 ± 0.00040 | 0.738 ± 0.082 |
| **Australia** | 0.00403 ± 0.00057 | 0.882 ± 0.047 |
| **South America** | 0.00244 ± 0.00015 | 0.769 ± 0.018 |
| **Asia+Europe** | 0.00314 ± 0.00027 | 0.862 ± 0.032 |
| **Asia+Europe+Australia** | 0.00380 ± 0.00026 | 0.894 ± 0.024 |

### 4.3.2 Haplotypes

A total of 47 haplotypes were identified from 314 individuals that consisted of 44 individuals from ASIA, 18 from Australia, 28 from Europe and 226 from South America (Table 4). Four most prevalent mtDNA COI haplotypes identified in this study were designated Harm_BC01, Harm_BC02, Harm_BC03, and Harm_BC04.

**Table 4:** Haplotypes distribution within individuals from Australia (AUS, n = 18), Asia (Asia, n = 44), Europe (EUR, n = 28), Argentina (ARG, n = 4), Uruguay (URY, n = 2), Paraguay (PRY, n = 10) and Brazilian states (Bahia BA, n = 96; Maranhão MA, n = 20; Mato Grosso MT, n = 19; Piauí PI, n = 39; Roraima RR, n = 14; Paraná PR, n = 3; Santa Catarina SC, n = 7; Rio Grande do Sul RS, n = 12).

| | | | | | | | **BRAZILIAN STATES** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Haplotypes** | **AUS** | **ASIA** | **EUR** | **ARG** | **URY** | **PRY** | **BA** | **MA** | **MT** | **PI** | **RR** | **PR** | **SC** | **RS** |
| **Harm_BC01** | 2 | 10 | 14 | | | 3 | 24 | 8 | 6 | 11 | 6 | 1 | 4 | 3 |
| **Harm_BC02** | | 4 | 2 | | 1 | | 36 | 9 | 6 | 18 | 5 | 1 | 2 | 3 |
| **Harm_BC03** | 1 | | 3 | 1 | | | 13 | 1 | 6 | 3 | 1 | | | |
| **Harm_BC04** | | 3 | 3 | 1 | | | 10 | | | 1 | | | | 4 |
| **Harm_BC05** | | | | | | | 5 | | | 2 | 1 | 1 | | |
| **Harm_BC06** | | 6 | 2 | | | | | 1 | | | | | | |
| **Harm_BC07** | | 1 | | | | | 2 | | | 3 | | | | |
| **Harm_BC08** | 5 | | | | | | | | | | | | | |
| **Harm_BC09** | | 5 | | | | | | | | | | | | |
| **Harm_BC10** | | 3 | 1 | | | | | | | | | | | |
| **Harm_BC11** | 3 | | | | | | | | | | | | | |

| | | | | | | | BRAZILIAN STATES | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Haplotypes** | **AUS** | **ASIA** | **EUR** | **ARG** | **URY** | **PRY** | **BA** | **MA** | **MT** | **PI** | **RR** | **PR** | **SC** | **RS** |
| **Harm_BC12** | 3 | | | | | | | | | | | | | |
| **Harm_BC13** | | | | | | 2 | | | | | | | 1 | |
| **Harm_BC14** | | | | | | | 2 | | | | | | | |
| **Harm_BC15** | | 2 | | | | | | | | | | | | |
| **Harm_BC16** | | | | 1 | | 1 | | | | | | | | |
| **Harm_BC17** | | | | | | 2 | | | | | | | | |
| **Harm_BC18** | 2 | | | | | | | | | | | | | |
| **Harm_BC19** | | 1 | | | | | | | | | | | | |
| **Harm_BC20** | | | 1 | | | | | | | | | | | |
| **Harm_BC21** | | | 1 | | | | | | | | | | | |
| **Harm_BC22** | | | 1 | | | | | | | | | | | |
| **Harm_BC23** | | | | | | | | | | 1 | | | | |
| **Harm_BC24** | | | | | | | | | | | 1 | | | |
| **Harm_BC25** | | 1 | | | | | | | | | | | | |
| **Harm_BC26** | | 1 | | | | | | | | | | | | |
| **Harm_BC27** | | 1 | | | | | | | | | | | | |
| **Harm_BC28** | | 1 | | | | | | | | | | | | |
| **Harm_BC29** | | 1 | | | | | | | | | | | | |
| **Harm_BC30** | | 1 | | | | | | | | | | | | |
| **Harm_BC31** | | 1 | | | | | | | | | | | | |
| **Harm_BC32** | | 1 | | | | | | | | | | | | |
| **Harm_BC33** | | 1 | | | | | | | | | | | | |
| **Harm_BC34** | | | | | | | 1 | | | | | | | |
| **Harm_BC35** | | | | | | | 1 | | | | | | | |
| **Harm_BC36** | | | | | | | 1 | | | | | | | |
| **Harm_BC37** | | | | | | | | 1 | | | | | | |
| **Harm_BC38** | | | | | | | | | 1 | | | | | |
| **Harm_BC39** | | | | | | | 1 | | | | | | | |
| **Harm_BC40** | 1 | | | | | | | | | | | | | |
| **Harm_BC41** | 1 | | | | | | | | | | | | | |
| **Harm_BC42** | | | | | | | | | | | | | | 1 |
| **Harm_BC43** | | | | | | 1 | | | | | | | | |
| **Harm_BC44** | | | | | 1 | | | | | | | | | |
| **Harm_BC45** | | | | | | | | | | | | | | 1 |
| **Harm_BC46** | | | | | | 1 | | | | | | | | |

| | | | | | | | BRAZILIAN STATES | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Haplotypes** | **AUS** | **ASIA** | **EUR** | **ARG** | **URY** | **PRY** | **BA** | **MA** | **MT** | **PI** | **RR** | **PR** | **SC** | **RS** |
| **Harm_BC47** | | | | 1 | | | | | | | | | | |

Brazil shares 5 mtDNA COI haplotypes with Asia (Harm_BC01, Harm_BC02, Harm_BC04, Harm_BC06 and Harm_BC07) and also 5 haplotypes with Europe (Harm_BC01, Harm_BC02, Harm_BC03, Harm_BC04 and Harm_BC06). Haplotypes Harm_BC01 and Harm_BC02 were present in all Brazilian states, and the Harm_BC01 haplotype was shared with all other locations, except Argentina and Uruguay. The haplotype Harm_BC02 in Brazil was shared with Asia, Europe and Uruguay. Brazil also had 12 unique haplotypes (Harm_BC05, Harm_BC14, Harm_BC23, Harm_BC24, Harm_BC34, Harm_BC35, Harm_BC36, Harm_BC37, Harm_BC38, Harm_BC39, Harm_BC42 and Harm_BC45), and 80% of Brazilian individuals belonged to haplotypes Harm_BC01, Harm_BC02, Harm_BC03, Harm_BC04 and Harm_BC05 (Figs. 2, 3 and 4).

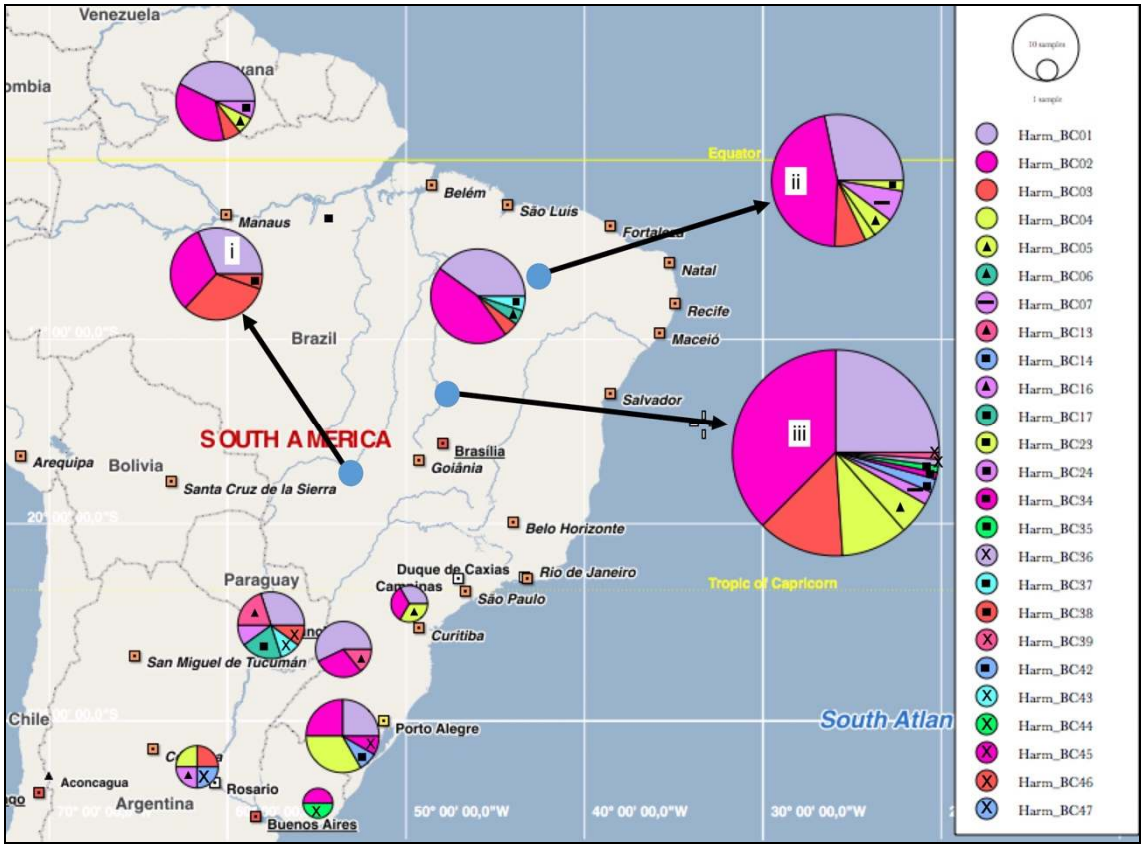


**Fig 2:** The haplotype distribution patterns and diversity of *Helicoverpa armigera* in South America using the partial mtDNA COI gene (548bp). Pie charts (i), (ii) and (iii) are from Mato Grosso, Piaui and Bahia states, respectively.

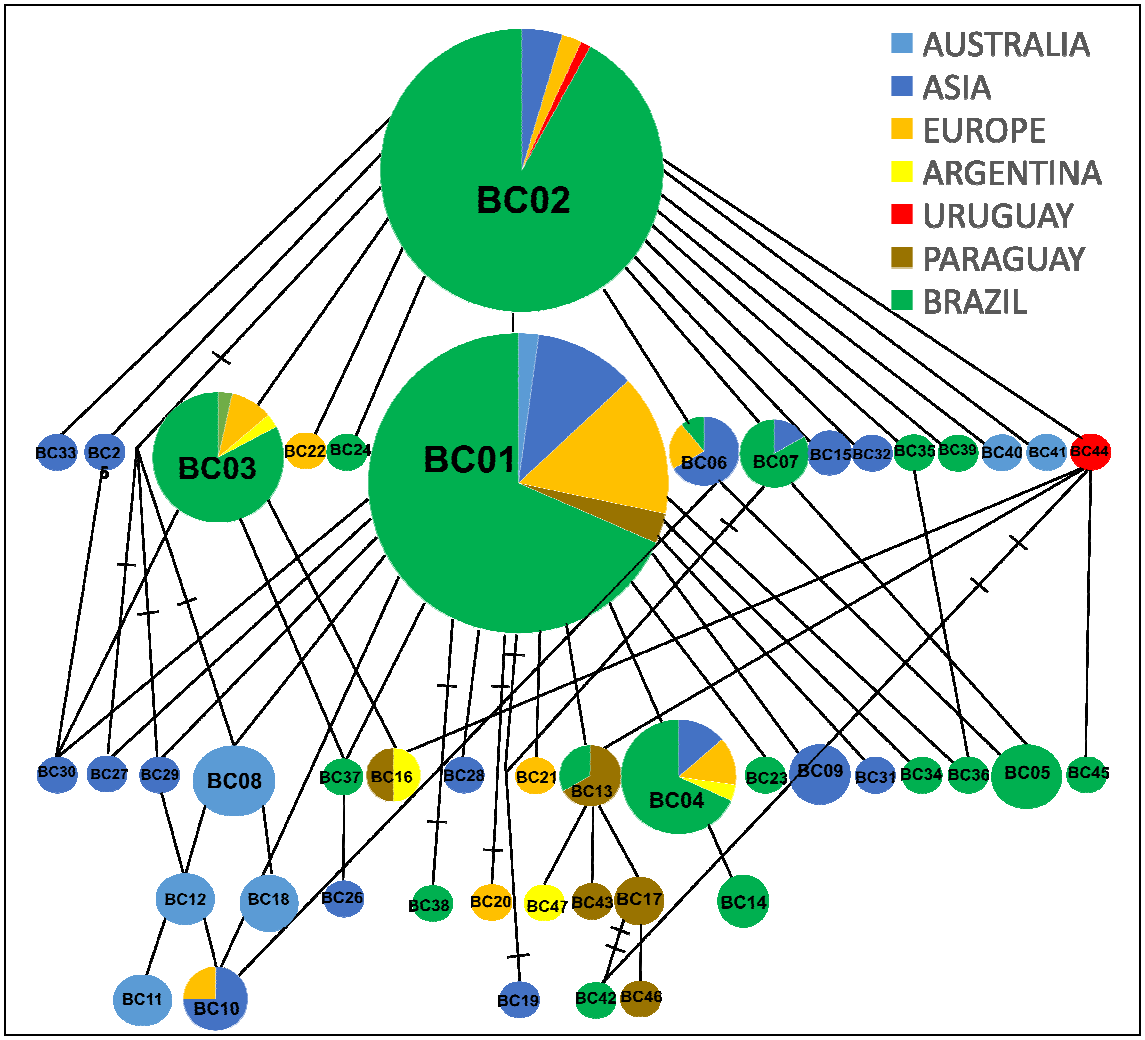


**Fig 3**: Haplotype network of *H. armigera* based on partial (548 bp) mtDNA COI gene, including samples from Australia, Asia, Europe, Argentina, Uruguay, Paraguay and Brazil. Each haplotype is represented by a circle, and is identified by 'Harm_BC01' to 'Harm_BC47'. Haplotype 'Harm_BC01' included 92 individuals; haplotype 'Harm_BC02' 87 individuals; haplotypes 'Harm_BC03', 'Harm_BC04', 'Harm_BC05' and 'Harm_BC06' have 29, 22, 9 and 9 individuals, respectively. All remaining haplotypes have less than 6 individuals each. Unique haplotypes are from 'Harm_BC19' to 'Harm_BC47'. Haplotypes are identified by haplotype numbers (i.e., BC##) in the figure. All haplotypes differed from each other by one base change, exceptions are between 'Harm_BC01'/'Harm_BC38', 'Harm_BC01'/'Harm_BC20', 'Harm_BC01'/'Harm_BC19', 'Harm_BC02'/'Harm_BC27', 'Harm_BC02'/'Harm_BC29', 'Harm_BC02'/'Harm_BC08', 'Harm_BC07'/'Harm_BC19', 'Harm_BC17'/'Harm_BC42', 'Harm_BC44'/'Harm_BC42' which differed from each other by two base changes.

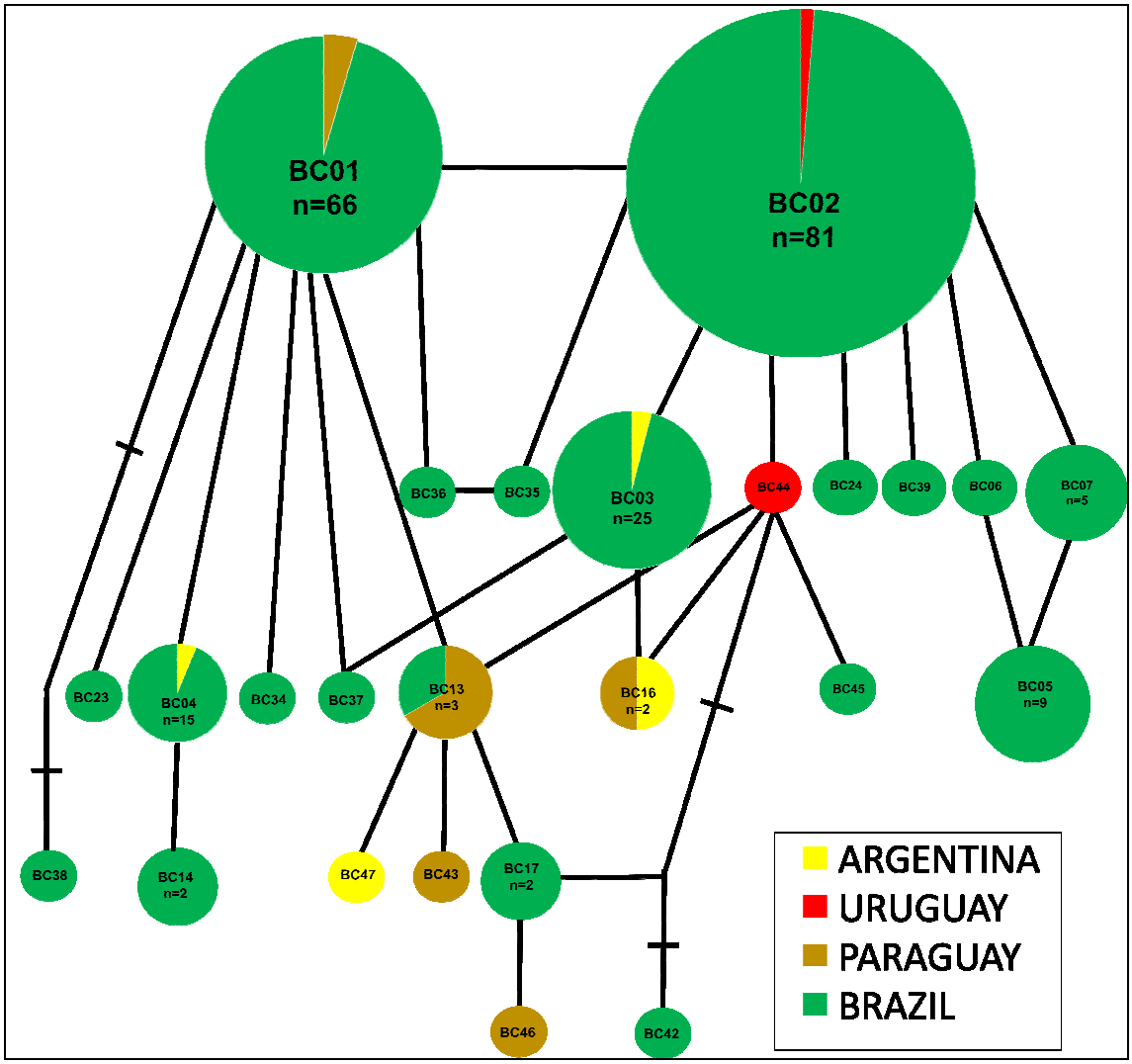


**Fig. 4:** Haplotype network of South American (Argentina, Uruguay, Paraguay, Brazil) *H. armigera* based on partial (548 bp) mtDNA COI gene. Haplotype Harm_BC01 included 66 individuals; haplotype Harm_BC02 has 81 individuals; haplotype Harm_BC03 has 25 individuals; Harm_BC04 has 15 individuals; Harm_BC05 has 9 individuals; Harm_BC07 has 5 individuals; Harm_BC13 has 3 individuals; Harm_BC14, Harm_BC16 and Harm_BC17 haplotypes each has two individuals. All remaining haplotypes have 1 individual each. Haplotypes are identified by their numbers in the figure. All haplotypes differed from each other by one base change, exceptions are between 'Harm_BC01'/'Harm_BC38' and 'Harm_BC44'/'Harm_BC42' which differed from each other by two base changes.

### 4.3.3 *Helicoverpa armigera* haplotypes distribution

A matrix table was constructed for the total recorded number of *H. armigera* individuals with particular mtDNA COI haplotypes across the 11 South American locations (i.e., Argentina, Uruguay, Paraguay, and Brazilian samples from various states: Bahia (BA), Maranhão (MA), Mato Grosso (MT), Piauí (PI), Roraima (RR), Paraná (PR), Santa Catarina (SC), and Rio Grande do Sul (RS) (Table 5).

**Table 5:** Detected mtDNA COI haplotypes within South America locations with the number of *Helicoverpa armigera* individuals carrying specific haplotypes at particular sites indicated.

| Haplotypes | ARG | URY | PRY | Brazilian states | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | BA | MA | MT | PI | RR | PR | SC | RS | n |
| **Harm_BC01** | 0 | 0 | 3 | 24 | 8 | 6 | 11 | 6 | 1 | 4 | 3 | 66 |
| **Harm_BC02** | 0 | 1 | 0 | 36 | 9 | 6 | 18 | 5 | 1 | 2 | 3 | 81 |
| **Harm_BC03** | 1 | 0 | 0 | 13 | 1 | 6 | 3 | 1 | 0 | 0 | 0 | 25 |
| **Harm_BC04** | 1 | 0 | 0 | 10 | 0 | 0 | 1 | 0 | 0 | 0 | 4 | 16 |
| **Harm_BC05** | 0 | 0 | 0 | 5 | 0 | 0 | 2 | 1 | 1 | 0 | 0 | 9 |
| **Harm_BC06** | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| **Harm_BC07** | 0 | 0 | 0 | 2 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 5 |
| **Harm_BC13** | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| **Harm_BC14** | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| **Harm_BC16** | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| **Harm_BC17** | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| **Harm_BC23** | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| **Harm_BC24** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| **Harm_BC34** | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| **Harm_BC35** | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| **Harm_BC36** | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| **Harm_BC37** | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| **Harm_BC38** | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| **Harm_BC39** | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| **Harm_BC42** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| **Harm_BC43** | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

| Haplotypes | ARG | URY | PRY | Brazilian states | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | BA | MA | MT | PI | RR | PR | SC | RS | n |
| **Harm_BC44** | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| **Harm_BC45** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| **Harm_BC46** | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| **Harm_BC47** | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | | | | | | | | | | | |
| **n** | 4 | 2 | 10 | 96 | 20 | 19 | 39 | 14 | 3 | 7 | 12 | 226 |

### 4.3.4 Validation of the null model

Results from validation of the null model indicated no tendency for the test to generate either Type I or Type II errors (i.e., having an almost perfect rectangular distribution of P-values across 10,000 statistical tests, with each test based on 10,000 random realisations of pseudo-observed data). At the 5% confidence level, almost exactly 5% of tests yielded a statistically significant results with randomly generated data (501/10,000 = 0.501 (Fig. 5a); 496/10,000 = 0.496 (Fig. 5b).

Fig. 5a

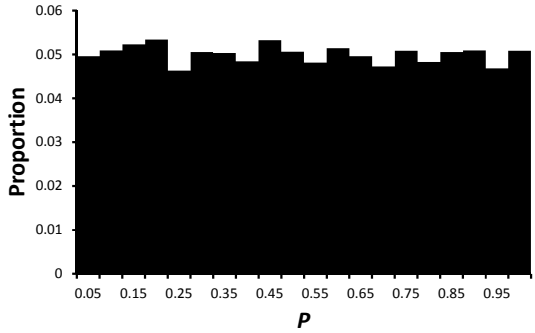


Fig. 5b

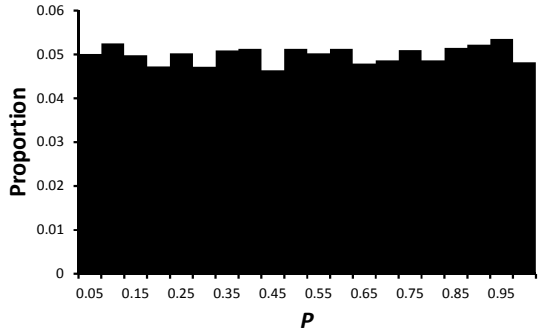


**Fig. 5**: Distribution of P-values from 10,000 significance tests using pseudo-observed data conforming to the null hypothesis. (a) Model validation results for matrices of size 25 rows x 11 columns, corresponding to the dimensions of the full dataset (Table 5). (b) Model validation results for matrices of size 25 rows x 2 columns, corresponding to the dimensions of the aggregated dataset for examining patterns between Brazil and non-Brazilian locations (see above).

### 4.3.5 Whole table analysis

Figure 6 shows the distribution of $\chi^2_{Ran}$under the null model. No random realisations were found to have a value of the tests statistic greater than or equal to that observed ($\chi^2_{Obs}$ = 719.2), thereby indicating highly significant non-randomness within the observations (P = 0.000), and strong support that at least one haplotype/location observation has observations that is either more or less than expected by chance alone.

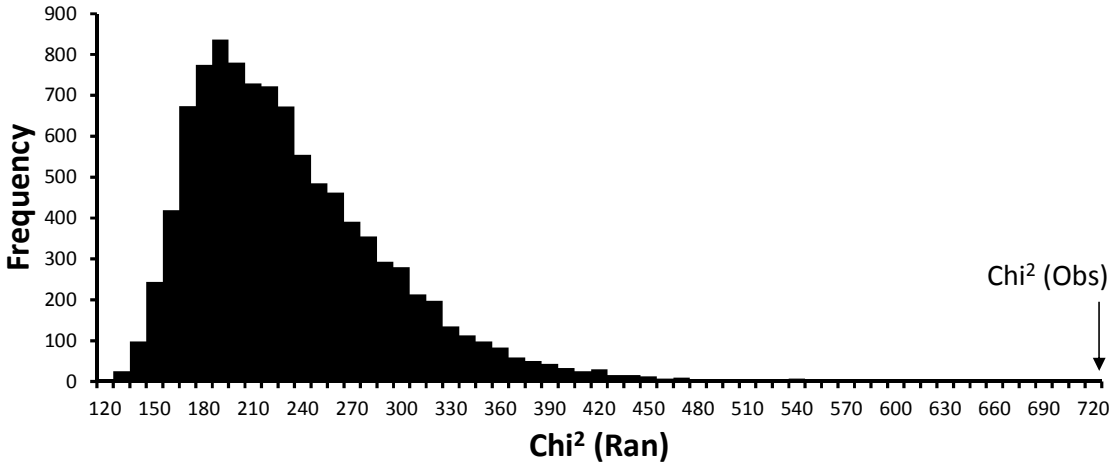


**Fig. 6:** Distribution of $\chi^2_{Ran}$ under the null model, and the location of $\chi^2_{Obs}$.

### 4.3.6 Sub-table analysis

Evidence from the whole table analysis strongly supported some mtDNA haplotypes were differentially distributed across sampling sites, and the $TS_{DIFF}$ analysis was therefore used to further identify those haplotypes that were unduly rare or common across the locations. The results indicated haplotype Harm_BC01 and Harm_BC02 were both simultaneously under-represented in Argentina (ARG), Paraguay (PRY) and in the Brazilian state of Bahia (BA), and overly represented in the

Brazilian states of Maranhão (MA), Piaui (PI), and Santa Catarina (SC) (Table 6). The analysis also found evidence to support haplotypes Harm_BC03, Harm_BC04 and Harm_BC07 as being sporadically overly represented in four Brazilian states (BA, MT, PI and RS) (dark blue cells on Table 6). Finally, there is also evidence that supported seven relatively uncommon haplotypes (i.e., Harm_BC13, Harm_BC16, Harm_BC17, Harm_BC43, Harm_BC44, Harm_BC46, and Harm_BC47) as being overly represented in the three neighbouring countries (Table 6).

**Table 6:** Table of $TS_{DIFF}$ values and indication of statistically significant deviations from the null model. Pale red cells are cells with a non-significantly lower number of observations than expected at random, and pale blue cells non-significantly higher numbers of observations. Dark red cells indicate significantly lower numbers of observations, with a P-value < 0.05. Dark blue cells indicate significantly higher numbers of observations, with a P-value < 0.05. Medium blue cells indicate marginally significant higher numbers of observations (P-values between 0.05 and 0.10).

| | Non-Brazil | | | Brazilian states | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **ARG** | **URY** | **PRY** | **BA** | **MA** | **MT** | **PI** | **RR** | **PR** | **SC** | **RS** |
| **Harm_BC01** | -1.54 | -0.58 | 0.07 | -4.21 | 2.03 | 0.15 | -0.19 | 1.90 | 0.24 | 1.99 | 0.14 |
| **Harm_BC02** | -1.37 | 0.28 | -4.02 | 1.49 | 1.59 | -0.85 | 4.15 | 0.00 | -0.11 | -0.41 | -0.74 |
| **Harm_BC03** | 0.65 | -0.22 | -1.05 | 2.06 | -1.26 | 3.90 | -1.39 | -0.49 | -0.29 | -0.68 | -1.22 |
| **Harm_BC04** | 0.78 | -0.14 | -0.58 | 2.97 | -1.44 | -1.34 | -1.83 | -0.92 | -0.20 | -0.44 | 3.14 |
| **Harm_BC05** | -0.13 | -0.08 | -0.34 | 1.06 | -0.69 | -0.67 | 0.35 | 0.46 | 0.87 | -0.29 | -0.55 |
| **Harm_BC06** | -0.01 | -0.01 | -0.04 | -0.40 | 0.92 | -0.08 | -0.18 | -0.07 | -0.02 | -0.04 | -0.07 |
| **Harm_BC07** | -0.07 | -0.04 | -0.19 | -0.23 | -0.36 | -0.36 | 2.13 | -0.31 | -0.08 | -0.17 | -0.33 |
| **Harm_BC13** | -0.04 | -0.03 | 1.89 | -1.38 | -0.21 | -0.21 | -0.49 | -0.18 | -0.05 | 0.90 | -0.20 |
| **Harm_BC14** | -0.03 | -0.02 | -0.08 | 1.23 | -0.16 | -0.16 | -0.37 | -0.14 | -0.04 | -0.08 | -0.15 |
| **Harm_BC16** | 0.97 | -0.02 | 0.92 | -0.77 | -0.16 | -0.16 | -0.37 | -0.14 | -0.04 | -0.08 | -0.15 |
| **Harm_BC17** | -0.03 | -0.02 | 1.92 | -0.77 | -0.16 | -0.16 | -0.37 | -0.14 | -0.04 | -0.08 | -0.15 |
| **Harm_BC23** | -0.01 | -0.01 | -0.04 | -0.38 | -0.08 | -0.08 | 0.81 | -0.07 | -0.02 | -0.04 | -0.08 |
| **Harm_BC24** | -0.01 | -0.01 | -0.04 | -0.38 | -0.08 | -0.08 | -0.19 | 0.93 | -0.02 | -0.04 | -0.08 |
| **Harm_BC34** | -0.01 | -0.01 | -0.04 | 0.62 | -0.08 | -0.08 | -0.19 | -0.07 | -0.02 | -0.04 | -0.08 |
| **Harm_BC35** | -0.01 | -0.01 | -0.04 | 0.61 | -0.08 | -0.08 | -0.18 | -0.07 | -0.02 | -0.04 | -0.08 |

| | Non-Brazil | | | Brazilian states | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **ARG** | **URY** | **PRY** | **BA** | **MA** | **MT** | **PI** | **RR** | **PR** | **SC** | **RS** |
| **Harm_BC36** | -0.01 | -0.01 | -0.04 | 0.62 | -0.08 | -0.08 | -0.19 | -0.07 | -0.02 | -0.04 | -0.08 |
| **Harm_BC37** | -0.01 | -0.01 | -0.04 | -0.38 | 0.92 | -0.08 | -0.18 | -0.07 | -0.02 | -0.04 | -0.08 |
| **Harm_BC38** | -0.01 | -0.01 | -0.04 | -0.38 | -0.08 | 0.92 | -0.19 | -0.07 | -0.02 | -0.04 | -0.08 |
| **Harm_BC39** | -0.01 | -0.01 | -0.04 | 0.62 | -0.08 | -0.08 | -0.19 | -0.07 | -0.02 | -0.04 | -0.08 |
| **Harm_BC42** | -0.01 | -0.01 | -0.04 | -0.38 | -0.08 | -0.08 | -0.18 | -0.07 | -0.02 | -0.04 | 0.91 |
| **Harm_BC43** | -0.01 | -0.01 | 0.96 | -0.39 | -0.08 | -0.08 | -0.18 | -0.06 | -0.02 | -0.04 | -0.10 |
| **Harm_BC44** | -0.01 | 0.99 | -0.04 | -0.38 | -0.08 | -0.08 | -0.17 | -0.06 | -0.01 | -0.04 | -0.11 |
| **Harm_BC45** | -0.01 | -0.01 | -0.04 | -0.38 | -0.08 | -0.08 | -0.16 | -0.06 | -0.01 | -0.03 | 0.85 |
| **Harm_BC46** | -0.01 | -0.01 | 0.96 | -0.33 | -0.07 | -0.06 | -0.11 | -0.03 | -0.01 | -0.02 | -0.32 |
| **Harm_BC47** | 1.00 | -0.00 | -0.01 | -0.13 | -0.05 | -0.05 | -0.16 | -0.11 | -0.04 | -0.11 | -0.32 |

**Note:** Because the row and column totals were fixed under the null model as previously detailed (see Material and Methods), there was therefore an inherent dependency amongst the $Z_{i,j}$ cells within the table, where the *P*-values for one cell calculated from $TS_{DIFF}$ dictated the *P*-values of other cells (i.e., an excess of observations in one table cell must be balanced by a deficit in another to maintain the row and column constraints). This dependency should be borne in mind when interpreting the patterns at the within-table scale.

### 4.3.7 Brazil vs. Non-Brazil

In the 'Brazil vs. non-Brazil' treatment of haplotype distribution data (Table 7), non-randomness of haplotype distribution within the matrix was again confirmed by the $\chi^2_{Obs}$ analysis ($\chi^2_{Obs}$= 164.0, P-value < 0.000; Fig. 7). The irregular distribution of Fig. 7 reflected a smaller data table and, therefore, fewer possible combinations of allowable observations to fulfil the row and column constraints.

**Table 7:** Observations of haplotypes within South American locations summed into two broad categories.

| Haplotypes | Non-Brazil | Brazil | n |
|---|---|---|---|
| Harm_BC01 | 3 | 63 | 66 |
| Harm_BC02 | 1 | 80 | 81 |
| Harm_BC03 | 1 | 24 | 25 |
| Harm_BC04 | 1 | 15 | 16 |
| Harm_BC05 | 0 | 9 | 9 |
| Harm_BC06 | 0 | 1 | 1 |
| Harm_BC07 | 0 | 5 | 5 |
| Harm_BC13 | 2 | 1 | 3 |
| Harm_BC14 | 0 | 2 | 2 |
| Harm_BC16 | 2 | 0 | 2 |
| Harm_BC17 | 2 | 0 | 2 |
| Harm_BC23 | 0 | 1 | 1 |
| Harm_BC24 | 0 | 1 | 1 |
| Harm_BC34 | 0 | 1 | 1 |
| Harm_BC35 | 0 | 1 | 1 |
| Harm_BC36 | 0 | 1 | 1 |
| Harm_BC37 | 0 | 1 | 1 |
| Harm_BC38 | 0 | 1 | 1 |
| Harm_BC39 | 0 | 1 | 1 |
| Harm_BC42 | 0 | 1 | 1 |
| Harm_BC43 | 1 | 0 | 1 |
| Harm_BC44 | 1 | 0 | 1 |
| Harm_BC45 | 0 | 1 | 1 |
| Harm_BC46 | 1 | 0 | 1 |
| Harm_BC47 | 1 | 0 | 1 |
| n | 16 | 210 | 226 |

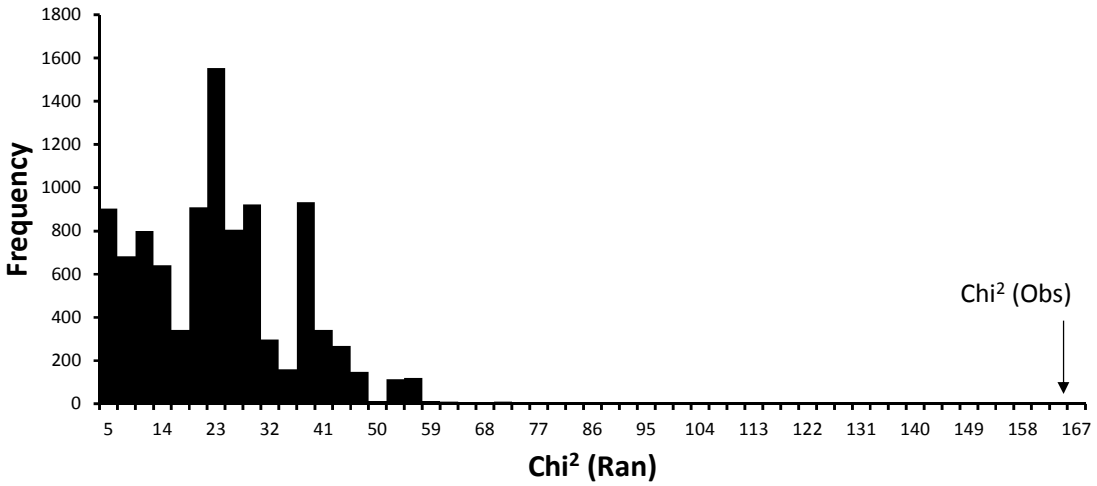


**Fig 7:** Distribution of $\chi^2_{Ran}$ under the null model, and the location of $\chi^2_{Obs}$ for the Brazil haplotypes vs non-Brazil haplotypes comparison.

Consistent with the full-table analysis (i.e., Table 6), at the scale of the individual haplotypes Harm_BC01 and Harm_BC02 were simultaneously under-represented in Non-Brazil and over-represented in Brazil. For haplotypes Harm_BC13, Harm_BC16 and Harm_BC17 the opposite was true. The over-representation of unique haplotypes Harm_BC43, Harm_BC44, Harm_BC46 and Harm_BC47 at Non-Brazil locations was also detected in the analysis but with P-values that just lie outside of the range usually considered significant (i.e., P-values in the range of 0.100 – 0.114).

**Table 8:** Table of $TS_{DIFF}$ for the Brazil vs Non-Brazil comparison of associated P-values.

| Haplotypes | Non-Brazil | | Brazil | |
|---|---|---|---|---|
| | $TS_{DIFF}$ | P-values | $TS_{DIFF}$ | P-values |
| Harm_BC01 | -1.90 | 0.000 | 1.90 | 0.000 |
| Harm_BC02 | -5.20 | 0.000 | 5.20 | 0.000 |
| Harm_BC03 | -0.70 | 0.895 | 0.70 | 0.895 |
| Harm_BC04 | -0.04 | 1.000 | 0.04 | 1.000 |
| Harm_BC05 | -0.50 | 1.000 | 0.50 | 1.000 |

| Haplotypes | Non-Brazil | | Brazil | |
|---|---|---|---|---|
| | $TS_{DIFF}$ | P-values | $TS_{DIFF}$ | P-values |
| Harm_BC06 | -0.06 | 1.000 | 0.06 | 1.000 |
| Harm_BC07 | -0.28 | 1.000 | 0.28 | 1.000 |
| Harm_BC13 | 1.83 | 0.014 | -1.83 | 0.014 |
| Harm_BC14 | -0.12 | 1.000 | 0.12 | 1.000 |
| Harm_BC16 | 1.89 | 0.005 | -1.89 | 0.005 |
| Harm_BC17 | 1.89 | 0.004 | -1.89 | 0.004 |
| Harm_BC23 | -0.06 | 1.000 | 0.06 | 1.000 |
| Harm_BC24 | -0.06 | 1.000 | 0.06 | 1.000 |
| Harm_BC34 | -0.06 | 1.000 | 0.06 | 1.000 |
| Harm_BC35 | -0.06 | 1.000 | 0.06 | 1.000 |
| Harm_BC36 | -0.06 | 1.000 | 0.06 | 1.000 |
| Harm_BC37 | -0.06 | 1.000 | 0.06 | 1.000 |
| Harm_BC38 | -0.06 | 1.000 | 0.06 | 1.000 |
| Harm_BC39 | -0.06 | 1.000 | 0.06 | 1.000 |
| Harm_BC42 | -0.06 | 1.000 | 0.06 | 1.000 |
| Harm_BC43 | 0.94 | 0.116 | -0.94 | 0.116 |
| Harm_BC44 | 0.94 | 0.115 | -0.94 | 0.115 |
| Harm_BC45 | -0.06 | 1.000 | 0.06 | 1.000 |
| Harm_BC46 | 0.94 | 0.113 | -0.94 | 0.113 |
| Harm_BC47 | 0.95 | 0.108 | -0.95 | 0.108 |

**Note:** With only two locations the values for $TS_{DIFF}$ are the same but of opposite magnitude, due to the row and column constraints in the null model. I.e. observing a significant deviation for one of the locations implies an equal and opposite deviation for the other.

## 4.4 Discussion

Statistical analyses of the haplotype distribution patterns supported the observation that some of the *H. armigera* haplotypes most commonly found in Brazil appeared to be uncommon in Argentina (i.e., Harm_BC01) and Paraguay (Harm_BC02), and conversely, that some of the less common and/or unique haplotypes found in the non-Brazilian countries (e.g., Harm_BC13, Harm_BC16, Harm_BC17, Harm_BC43,

Harm_BC44, Harm_BC46, and Harm_BC47) appeared disproportionately uncommon in Brazil. Within individual countries, disproportionally over-represented haplotypes were identified (e.g., see Table 5), and within Brazil, the rare Harm_BC13 haplotype was detected in Santa Catarina which neighbours Paraguay (moth samples with the Harm_BC13 haplotype were collected at sites separated by approximately 350 km).

Population structure studies in *H. armigera* based on the mtDNA genes have found a general lack of substructure even for populations separated by considerable geographic distances (e.g., Spackman and McKechnie 1995; Behere et al. 2007), and is also supported by studies based on nuclear markers (e.g., Daly and Gregg 1985; Nibouche et al. 1998; Zhou et al. 2000; Xiao-feng et al. 2000; Endersby et al. 2007; Vassal et al. 2008; Behere et al. 2013). In Brazil, gene flow patterns of *H. armigera* showed non-significant levels of population substructure as indicated by the low and non-significant $F_{ST}$ values (i.e., having substantial levels of gene flow between populations) (Mastrangelo et al. 2014; Leite et al. 2014). Furthermore, the majority of studied Brazilian *H. armigera* individuals also have the most common haplotypes (i.e., Harm_BC01, Harm_BC02). In this chapter, the spatial haplotype distribution patterns presented have greatly relied on the mtDNA COI haplotype survey efforts of both Mastrangelo et al. (2014) and Leite et al. (2014), and a repeat of population structure/gene flow analysis would unlikely lead to a different conclusion even if the limited Brazilian individuals sampled from PR, RS and SC states were included (i.e., any potential population structure signal is likely to be masked by the statistically significant over-representation of the most common Harm_BC01 and Harm_BC02 haplotypes (e.g., see Tables 5 and 6).

Given that sufficient gene flow had been detected in Brazilian *H. armigera* populations leading to non-significant population substructure, and that positive detection of *H. armigera* in Argentina and Paraguay were also confirmed (Senave 2013; Murúa et al. 2014) post incursions in Brazil, it would be intuitive to assume that this likely represent effects of 'natural' migration, and/or movements of contaminated agricultural commodities between Brazil (i.e., which may act as source populations) and other regions of the New World. It was therefore surprising to find in the limited field-caught non-Brazilian *H. armigera* individuals, the presence of unique mtDNA haplotypes as yet to be reported in Brazil. Equally as unexpected was the non-detection of most common haplotypes such as Harm_BC01 and/or Harm_BC02 in these countries (Tables 5-8), given that they represented the major haplotypes in Brazil.

To further explain the observed heterogeneous haplotype distribution patterns across the Cone Sul regions as compared with 'the rest of Brazil', two hypotheses may be put forward: (I) intrinsic factors associated with new biological incursions (e.g., stochastic lineage sorting, survival/reproductive variability, etc.) in a new environment, and (II) different (and possibly independent) incursion events of *H. armigera* into Brazil and non-Brazil countries.

In Brazil where the incursion of *H. armigera* was first reported, stochastic lineage sorting of founding populations (i.e., hypothesis I), potentially pointing to lower population density and variability at population introduction phase and/or lag-phase, high variability of reproductive success rates (e.g., Gaither et al. 2013), and variable adaptation success rates (e.g., differential response to attacks by parasitoids, predation rates, susceptibility to viral/bacterial/fungal pathogen attacks, climatic

stress, etc.) to the novel New World environments at the early incursion stages. In Brazil, that significant populations were detected (Leite et al. 2014; Mastrangelo et a. 2014) and with the unexpectedly high genetic diversity as represented by multiple maternal lineages (e.g., Fig. 2) would likely indicate a complex incursion history of *H. armigera* into Brazil as previously hypothesised (e.g., Tay et al. 2013). For example, repeated incursions/releases in different Brazilian sites could function as population source of invaders that helped increase propagule pressure and raise and maintain diversity which is important in sustaining an incipient population (Lockwood 2005; Simberloff 2009; Chown et al. 2014).

These suggested scenarios involving the highly volatile and variable periods of an exotic organism's invasive biology, such as during the initial population introduction, lag and growth phases contrasted the proposed scenarios offered by Leite et al. (2014), where repeated bottleneck effects such as potentially associated with differential pest control/management strategies were deemed most likely factors that underpinned the rapid population expansion signatures in both *H. armigera* and the New World endemic and closely related *H. zea*.

Repeated introductions and high propagule pressure are increasingly being recognised as important factors that underpinned establishment success of an alien species (Simberloff 2009; Chown et al. 2014). With repeated introduction events, the likelihood of diverse maternal lineages that ultimately contribute to propagule pressure is high, especially in global invasive pests such as *H. armigera*. Together with lineage sorting and stochastic processes (e.g., demographic, environmental; Simberloff 2009) being experienced by the invasive species in the new environment, sampling of the mtDNA COI gene and the construction of a haplotype network will likely appear

similar to one of rapid population expansion (i.e., a 'star-shaped' haplotype network). This scenario, of a 'star-shaped' haplotype network, as detected in *H. armigera* populations in the South Americas (e.g., Fig. 3 of this study; Leite et al. 2014 (Fig. 2)), differed fundamentally to that reported for *H. zea* (e.g., Behere et al. 2007 (Fig. 1); Leite et al. 2014 (Fig 2)), where the *H. zea* founder population (arrived in the New Word at *ca*. 1.5 million years ago; Behere et al. 2007), was able to undergo population expansion.

Multiple introduction history of an invasive pest insect that resulted in mtDNA genetic signature similar to a rapid population expansion signature, has also been reported in the Asian citrus psyllids *Diaphorina citri* (Hemiptera: Liviidae) by Guidolin et al. (2014) in Brazil. Although unique *H. armigera* mtDNA COI haplotypes can result at anytime post incursion into Brazil (i.e., as a result of rapid population expansion in a new environment; Leite et al. 2014; Mastrangelo et al. 2014), its occurrence must nevertheless be regarded as of low probability (e.g., based on a mtDNA COI divergence rate of 2.69% per million year in insect (Papadopoulou et al. 2010); approximately 6,000 years post incursion (equivalent to *ca*. 60.000-66.000 generations assuming an upper limit of 10-11 generations/year in *H. armigera* in the tropics (Fitt 1989)) would be needed for a star-shaped haplotype network to occur assuming a single *H. armigera* founder.

With the migration and dispersal ability of *H. armigera* in mind, the high frequency of Harm_BC01 and Harm_BC02 haplotypes in Brazil populations, and that previously studies have indicated a lack of population structures in Brazil, it was perhaps unexpected to observe statistically significant spatial mtDNA COI haplotype patterns in the Cone Sul region. For example, haplotypes Harm_BC13 and Harm_BC17

for Paraguay and Harm_BC44 and Harm_BC47 for Uruguay and Argentina, respectively, were over-represented in the respective countries (and to a lesser extent, also the over-representation of haplotypes Harm_BC43, Harm_BC44, Harm_BC46 and Harm_BC47, although the P-values (0.100 – 0.114) lie outside of the range usually considered significant), thereby adding support that these maternal lineages likely originated from non-Brazil source populations (i.e., hypothesis II).

Presence of *H. armigera* in Argentina was only recently confirmed by Murúa et al. (2014) and by Arneodo et al. (2015) based on morphological characters and the mtDNA COI gene, respectively. Presence of *H. armigera* in Paraguay has also been reported based on morphological characters (SENAVE 2013), while in Uruguay, positive confirmation of *H. armigera* based on the partial mtDNA COI gene represents the first time this insect is recorded in that country. Although the limited sample size of this study rendered a genetic diversity survey of *H. armigera* in these non-Brazilian countries as preliminary, the results nevertheless provided the first insights into potential patterns of biological incursion in the Cone Sul region of South America. Efforts must now be made to re-examine biosecurity protocols relating to phytosanitary practices of agricultural and horticultural commodities that are entering Argentina, Paraguay, and Uruguay. That the statistically significant over-representation of unique haplotypes was not shared between Brazil and her neighbouring southern countries, added support to the hypothesis of multiple independent incursions of *H. armigera* in South America.

Findings that populations of *H. armigera* in the South American continent likely represented multiple independent incursion events will have significant implications to pest and resistance management strategies in the New World. For example,

populations of *H. armigera* around the world have developed resistance to conventional pesticides (e.g., Yang et al. 2007; Achaleke et al. 2010; Zhang et al. 2012; Nair et al. 2013), and it would be important to know whether South American populations have also arrived with these resistance mechanisms. Increasing genetic diversity is a key factor that underpins increasing invasion success (Wares et al. 2005; Ellestrand and Schierenbeck 2000). While significant levels of genetic diversity now exist in Brazil, and propagule pressure has also therefore decreased, however that the genetic make-up of this pest insect could be further bolstered by likely unrelated source populations from other parts of the Old World will further complicate and challenge management strategies. As pointed out by De Barro et al. (2011), measures to restrict the recruitment of additional genetic diversity should be maintained even after establishment and spread has occurred, so as to avoid increasing the genetic diversity of damaging invasive pests.

**4.5 References**

Achaleke J, Brevault T. 2010. Inheritance and stability of pyrethroid resistance in the cotton bollworm *Helicoverpa armigera* (Lepidoptera: Noctuidae) in Central Africa. *Pest Management Science* 66: 137-141.

Albernaz KC, Silva-Brandao KL, Fresia P, Consoli FL, Omoto C. 2012. Genetic variability and demographic history of *Heliothis virescens* (Lepidoptera: Noctuidae) populations from Brazil inferred by mtDNA sequences. *Bulletin of Entomological Research* 102: 333-343.

Arnemann JA, Guedes JVC, Stacke RF, Melo, A, Perini CR, Curioletti, LE. 2014. *Helicoverpa armigera*: até no inverno. *Cultivar Grandes Culturas* 183: 26-28.

Arneodo JD, Balbi EI, Flores FM, Sciocco-Cap A. 2015. Molecular identification of *Helicoverpa armigera* (Lepidoptera: Noctuidae: Heliothinae) in Argentina and development of a novel PCR-RFLP method for its rapid differentiation from *H. zea* and *H. gelotopoeon. Journal of Economic Entomology.* DOI: http://dx.doi.org/10.1093/jee/tov254.

Behere GT, 2008. Genetic variation in field populations of the cotton bollworm, *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae). PhD Thesis, Department of Genetics, The University of Melbourne, Australia. XV+159pp.

Behere GT, Tay WT, Russell DA, Heckel DG, Appleton BR, Kranthi KR, Batterham P. 2007. Mitochondrial DNA analysis of field populations of *Helicoverpa armigera* (Lepidoptera: Noctuidae) and of its relationship to *H. zea*. *BMC Evolutionary Biology* 7: 117-127.

Behere GT, Tay WT, Russell DA, Kranthi KR, Batterham P. 2013. Population genetic structure of the cotton bollworm *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae) in India as inferred from EPIC-PCR DNA markers. *PLoS ONE* 8(1): e53448.

Bounfour M, Jebbour F, Wadjinny J, 2005. Biological traits of invasive insect species harmful to Moroccan agriculture. In: Plant protection and plant health in Europe: introduction and spread of invasive species. BCPC Conference, Humbolt University (DE) 95-100.

Buthelezi NM, Conlong DE, Zharare GE. 2012. The groundnut leaf miner collected from South Africa is identified by mtDNA CO1 gene analysis as the Australian soybean moth (*Aproaerema simplixella*) (Walker) (Lepidoptera: Gelechiidae). *African Journal of Agricultural Research* 7: 5285-5292.

Chown SL, Hodgins KA, Griffin PC, Oakeshott JG, Byrne M, Hoffmann AA. 2015. Biological invasions, climate change and genomics. *Evolutionary Applications* 8: 23-46. doi:10.1111/eva.12234.

Czepak C, Albernaz KC, Vivan LM, Guimaraes HO, Carvalhais T. 2013. Primeiro registro de ocorrência de *Helicoverpa armigera* (Hübner) Lepidoptera: Noctuidae) no Brasil. *Pesquisa Agropecuária Tropical* 43: 110-113.

Daly JC, Gregg P. 1985. Genetic variation in *Heliothis* in Australia: species identification and gene flow in the two pest species *H. armigera* (Hübner) and *H. punctigera* Wallengren (Lepidoptera: Noctuidae). *Bulletin of Entomological Research* 75: 169-184.

De Barro PJ, Liu S-S, Boykin LM, Dinsdale AB. 2011. *Bemisia tabaci*: A statement of species status. *Annual Review of Entomology* 56: 1-19.

Bueno A, Sosa-Gómez DR. 2014. The Old World bollworm in the neotropical region: The experience of Brazilian growers with *Helicoverpa armigera. Outlooks on Pest Management* 25: 261-264.

Ellestrand NC, Schierenbeck KA. 2000. Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences of the United States of America* 97: 7043-7050.

Endersby NM, Hoffmann AA, McKechnie SW, Weeks AR. 2007. Is there genetic structure in populations of *Helicoverpa armigera* from Australia? *Entomologia Experimentalis Applicata* 122: 253-263.

EPPO. 2015. EPPO Global Database (available online). [cited 05 September 2015]. Available from: < https://gd.eppo.int>.

Fitt GP. 1989. The ecology of *Heliothis* species in relation to agroecosystems. *Annual Review of Entomology* 34: 17-52.

Gaither MR, Bowen BW, Toonen RJ. 2013. Population structure in the native range predicts the spread of introduced marine species. *Proceedings of the Royal Society B,* 280: 20130409. http://dx.doi.org/10.1098/rspb.2013.0409

Gotelli NJ. 2000. Null Model Analysis of species co-occurrence patterns. *Ecology* 81: 2606-2621.

Guedes JVC, Arnemann JA, Perini CR, Melo AA, Röhrig A. Stacke RF, Machado MRR. 2014. *Helicoverpa armigera*: da invasão ao manejo na soja. *Revista Plantio Direto*, 137: 24-35.

Guidolin AS, Fresia P, Cônsoli FL. 2014. The genetic structure of an invasive pest, the Asian Citrus Psyllid *Diaphorina citri* (Hemiptera: Liviidae). *PLoS ONE* 9(12): e115749. doi:10.1371/journal.pone.0115749.

Hardwick DF. 1965. The corn earworm complex. *Memoirs of the Entomological Society of Canada* 97: 5-247.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. 2004. Identification of birds through DNA barcodes. *PLOS Biology* 2(10): e312. doi:10.1371/journal.pbio.0020312.

IBGE. 2015. Instituto Brasileiro de Geografia e Estatitica. Noticias. [cited 05 September 2015]. Available from: <http://www.ibge.gov.br/home/geociencias/geografia/default.shtm>.

Ivkosic SA, Gorman J, Lemic D, Mikac KM. 2014. Genetic monitoring of western corn rootworm (Coleoptera: Chrysomelidae) populations on a microgeographic scale. *Environmental Entomology* 43: 804-818.

King ABS. 1994. *Heliothis/Helicoverpa* (Lepidoptera: Noctuidae), pp. 39-106. In GA Matthews, JP Tunstall (eds.), Insect Pests of Cotton. CAB International, Wallingford.

Kiritani K, Yamamura K. 2003. Exotic insects and their pathways for invasion. In: Invasive species: vectors and management strategies. Ed. by Ruiz GM, Carlton JT, Island, Washington, DC 44-67.

Kriticos DJ, Ota N, Hutchinson WD, Beddow J, Walsh T, Tay WT, Borchert DM, Paula-Moreas SV, Czepak C, Zalucki MP. 2015. The potential distribution of invading *Helicoverpa armigera* in North America: Is it just a matter of time? *PLoS ONE* 10(7): e0133224.

Leite NA, Alves-Pereira A, Corrêa AS, Zucchi MI, Omoto C. 2014. Demographics and genetic variability of the New World bollworm (*Helicoverpa zea*) and the Old World bollworm (*Helicoverpa armigera*) in Brazil. *PLoS ONE* 9(11): e113286.

Librado P, Rozas J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452.

Lockwood JL, Casse P, Blackburn T. 2005 The role of propagule pressure in explaining species invasions. *Trends in Ecology & Evolution* 20: 223-228. (doi:10.1016/j.tree.2005. 02.004).

Lopes-da-Silva M, Sanches MM, Stancioli AR, Alves G, Sugayama R. 2014. The role of natural and human-mediated pathways for invasive agricultural pests: A historical analysis of cases from Brazil. *Agricultural Sciences* 5: 634-646.

Mack RN, Erneberg M. 2002. The United States naturalized flora: largely the product of deliberate introductions. *Annals of the Missouri Botanical Garden* 89: 176-189.

Manly BFJ. 2004. One-sided tests of bioequivalence with nonnormal distributions and unequal variance. *Journal of Agricultural, Biological, and Environmental Statistics* 9: 270-283.

Mastrangelo T, Paulo DF, Bergamo LW, Morais EGF, Silva M, Bezerra-Silva G, Azeredo-Espin AML. 2014. Detection and genetic diversity of a heliothine invader

(Lepidoptera: Noctuidae) from north and northeast of Brazil. *Journal of Economic Entomology* 107: 970-980.

Murúa MG, Scalora FS, Navarro FR, Cazado LE, Casmuz A, Villagrán ME, Lobos E, Gastaminza G. 2014. First record of *Helicoverpa armigera* (Lepidoptera: Noctuidae) in Argentina. *Florida Entomologist* 7: 854-856.

Nair R, Kalia V, Aggarwal KK, Gujar GT. 2013. Variation in the cadherin gene sequence of Cry1Ac susceptible and resistant *Helicoverpa armigera* (Lepidoptera: Noctuidae) and the identification of mutant alleles in resistant strains. *Curr Sci* 104: 215.

Nei M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.

Nibouche S, Bues R, Toubon JF, Poitout S. 1998. Allozyme polymorphism in the cotton bollworm *Helicoverpa armigera* (Lepidoptera: Noctuidae): comparison of African and European populations. *Heredity* 80: 438-445.

Oliveira CM, Auad AM, Mendes SM, Frizzas MR. 2013. Economic impact of exotic insect pests in Brazilian agriculture. *Journal of Applied Entomology* 137: 1-15.

Papadopoulou A, Anastasiou I, Vogler AP. 2010. Revisiting the insect mitochondrial molecular clock: The mid-Aegean trench calibration. *Molecular Biology and Evolution* 27: 1659-1672.

Pomari-Fernandes A, de Freitas Bueno A, Sosa-Gómez DR. 2015. *Helicoverpa armigera*: current status and future perspectives in Brazil. *Current Agricultural Science and Technology* 21: 1-7.

Schoener TW, Schoener A. 1983. The time to extinction of a colonizing propagule of lizards increases with island area. *Nature* 302: 332-334.

Senave. 2013. Senave en alerta tras ingreso de peligrosa plaga agrícola. [cited 05 September 2015]. Available from: <http://www.abc.com.py/edicion-impresa/economia/senave-en-alerta-tras-ingreso-de-peligrosa-plaga-agricola-629240.html>.

Simberloff D. 2009. The role of propagule pressure in biological invasions. *Annual Review of Ecology, Evolution, and Systematics* 40: 81-102.

Spackman ME, McKechnie SW. 1995. Assessing the value of mitochondrial DNA variation for detecting population subdivision in the cotton bollworm, *Helicoverpa armigera* (Lepidoptera: Noctuidae), in Australia. *Proceedings of the Beltwide Cotton Conference* 2: 811-813.

Specht A, Sosa-Gómez DR, de Paula-Moraes SV, Yano SAC. 2013. Identificação morfológica e molecular de *Helicoverpa armigera* (Lepidoptera: Noctuidae) e ampliação de seu registro de ocorrência no Brasil. *Pesquisa Agropecuária Brasileira* 48: 689-692.

Tamura K, Nei M, Kumar S. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences* 101: 11030-11035.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* 30: 2725-2729.

Tay WT, Soria MF, Walsh T, Thomazoni D, Silvie P, Behere GT, Anderson C, Downes S. 2013. A brave new world for an Old World pest: *Helicoverpa armigera* (Lepidoptera: Noctuidae) in Brazil. *PLoS ONE* 8(11): e80134.

USDA/APHIS (Animal and Plant Health Inspection Service). Detection of Old World Bollworm (*Helicoverpa armigera*) in Florida - FOR INFORMATION AND ACTION DA-2015-43 July 30, 2015. [cited 05 September 2015]. Available from: <http://www.aphis.usda.gov/plant_health/plant_pest_info/owb/downloads/DA-2015-43.pdf>.

Vassal JM, Brevault T, Achaleke J, Menozzi P. 2008. Genetic structure of the polyphagous pest *Helicoverpa armigera* (Lepidoptera: Noctuidae) across the Sub-Saharan cotton belt. *Communications in agricultural and applied biological sciences* 73: 433-437.

Wares JP, Hughes AR, Grosberg RK. 2005. Mechanisms that drive evolutionary change: insights from species introductions and invasions. In: Sax DF, Stachowicz JJ, Gaines SD, eds. Species invasions: Insights into ecology, evolution, and biogeography. Sinauer Associates: Massachusetts. pp 229-257.

Xiao-feng C, Sheng-jiang T, Ren-yi L, Ying W, Diamo L. 2000. Study on the genetic variation of the cotton bollworm *Helicoverpa armigera* (Hübner) populations in china. *Insect Science* 7: 243–249.

Yang Y, Chen H, Wu Y, Yang Y Wu S. 2007. Mutated cadherin alleles from a field population of *Helicoverpa armigera* confer resistance to *Bacillus thuringiensis* toxin Cry1Ac. *Applied and Environmental Microbiology* 73: 6939-6944.

Zhang H, Wu S, Yang Y, Tabashnik BE, Wu Y. 2012. Non-recessive Bt toxin resistance conferred by an intracellular cadherin mutation in field-selected populations of cotton bollworm. *PLoS One* 7: 12 e53418.

Zhou X, Faktor O, Applebaum SW, Coll M. 2000. Population structure of the pestiferous moth *Helicoverpa armigera* in the eastern Mediterranean using RAPD analysis. *Heredity* (Edinb). 85: 251-256.

# CHAPTER V

## GENERAL CONCLUSIONS AND FUTURE PERSPECTIVES

The present work explored the potential use of molecular tools to identify and study population genetics of two invasive pests in Brazil. The aims were: (1) to developed molecular markers, based on taxonomically well-supported pest species identity, for early and rapid species confirmation, and (2) to investigate the genetic diversity of invasive pest populations within an evolutionary biology framework to impact on biosecurity considerations. Two polyphagous and highly invasive insect species were examined in this study, and were chosen based on their economic impact, one with soybean as its major plant host, and the second as an ultimate polyphagous pest of all major food and fibre crops.

The first target species of Chapter II was the Soybean Stem Fly (SSF) *Melanagromyza sojae*, which was found damaging soybean (*Glycine max* L.) fields in southern Brazil. For this study, the species identity of this *Melanagromyza* species was first confirmed as *M. sojae* based on larval morphological characters with the assistance of Mr Hugh Brier, Senior Entomologist at the Queensland Department of Agriculture and Fisheries (Queensland, Australia), followed by molecular characterization of the fly's complete mitochondrial DNA (mtDNA) genome at the Commonwealth Scientific and Industry Research Organization (CSIRO) in Canberra, Australia.

Within the most important results from Chapter II are:

(i) confirmation of *M. sojae* by larval morphological characters, and by the mtDNA COI marker when compared with Australian SSF samples;

(ii) the first annotation of complete mitochondrial genomes of this significant *G. max* pest specie; and

(iii) the development of first DNA markers to survey population genetic diversity and for future studies of gene flow patterns in *M. sojae*.

The mtDNA genomes of three *M. sojae* were sequenced using the Illumina MiSeq NGS platform. Assembly and annotation of these mitochondrial genomes enabled the annotation of all 37 mtDNA genes (13 protein coding genes, 22 tRNAs, 2 rRNAs), and also the identification of the putative replication origin (i.e, the A+T-rich) region, with estimation of complete mitochondrial genome sequence length, base substitution patterns, and nucleotide polymorphism rates. Molecular characterisation of the SSF mitochondrial DNA genomes also enabled robust species-specific PCR markers to be developed for three mtDNA genes, targeting the mtDNA COI (for generating dataset comparable with the iBoL DNA database), ATP8/ATP6 and ND4 genes (regions identified as having the highest nucleotide diversity in the SSF mitochondrial DNA genomes (excluding the A+T-rich region) by a sliding window analysis). A partial mtDNA COI Maximum Likelihood (ML) phylogeny was also constructed to help infer the phylogenetic position of the Brazilian SSF among previously sequenced Agromyzidae flies, with the result indicating likely basal position for *M. sojae* as compared with other mtDNA COI sequence characterized *Melanagromyza* species.

With the occurrence of *M. sojae* in Rio Grande do Sul and Santa Catarina states confirmed, Chapter III aimed to understand the population genetic diversity of SSF using the newly developed SSF-specific mtDNA COI PCR marker (Chapter II), and discuss potential biosecurity implications of incursions by *M. sojae* in Brazil.

Highlights of results from Chapter III are:

(i) the unexpectedly high nucleotide diversity in the modest *M. sojae* sample size from Santa Catarina and Rio Grande do Sul;

(ii) the high number of mtDNA haplotypes identified suggested populations in Brazil potentially involved multiple incursions, although it was currently also not possible to differentiate from a single mixed population founder event; and

(iii) first insights into the biosecurity, agricultural and socio-economic implications of *M. sojae* incursions in Brazil, and the significant genomic resources and research opportunities into SSF metagenomics from NGS data generated that is currently awaiting detailed future studies.

In Chapter IV the mtDNA COI partial gene of *H. armigera* specimens collected from Uruguay, Argentina, Paraguay, and also from three southern Brazilian states (Parana (PR); Rio Grande do Sul (RS), Santa Catarina (SC)) were analysed with the aims to better characterise the genetic diversity of this highly invasive pest in the Cone Sul regions, and to explore potential new biosecurity concerns using population-wide genetic diversity signatures.

The main results from this study are:

(i) confirmation, at the molecular level, that *H. armigera* is present in soybean fields in Uruguay and Paraguay, and also in RS, SC, and PR;

(ii) first molecular genetic diversity characterisation of *H. armigera* in the Cone Sul region (Paraguay, Uruguay, Argentina, and southern Brazilian states of RS, SC and PR);

(iii) detection of unexpected and unique haplotype diversity in *H. armigera* populations from Uruguay, Argentina and Paraguay; and

(iv) findings that populations of *H. armigera* in the South American continent likely represented multiple independent incursion events, thereby highlighting potential biosecurity concerns in the Cone Sul territory.

On the occurrence of SSF in Brazil, any estimates of losses to Brazilian soybean fields caused by this insect pest would be premature and presumptive, as knowledge of how widespread the species is across Brazil's soybean cropping areas, and of its potential presence in neighbouring countries (e.g., Paraguay,) would first be needed. Furthermore, there is also a lack of knowledge with regards how this pest affects soybean cultivars, and their responses to climatic and environmental determinants, and to effects of biological and chemical control practices in Brazil. To better able to infer incursion patterns and the origin of SSF in Brazil, global surveying of population genetic diversity, combining the use of molecular genetic markers such as have been developed in this study (i.e., mtDNA markers) and nuclear DNA markers (e.g., microsatellite DNA markers; EPIC-PCR markers; whole genome SNPs as obtained using NGS methods, i.e., Genotyping-by-sequencing (GBS)) will be recommended.

The biosecurity concerns in the Cone Sul territory is further highlighted by the findings that populations of *H. armigera* likely represented multiple independent incursion events.

Significant implications/challenges to *H. armigera* resistance management strategies for these countries may also be expected, as populations of *H. armigera* around the world have developed resistance to conventional pesticides. The next

studies should involve whether South American populations have also arrived with these resistance mechanisms. A good example with regard to the management of insecticide resistance is one that is being applied by the Australian cotton industry with significant scientific input from researchers, and relates to active bioassay testing of field collected samples for monitoring of resistance alleles frequencies in *Helicoverpa* species to Bt Cry1Ac and Bt Cry2Ab. with findings translated into management policies that are actively adopted by growers. Monitoring of genes that underpin resistance to insecticides in South American populations of *H. armigera* will be essential to future management of this pest for long term benefit to the agriculture industry sector, especially those that are relying on Bt GM technology.

The most important biosecurity implications to the Cone Sul countries relates to the need to reassess individual countries' border protection and quarantine inspection protocols for the detection of phytosanitary pests and pathogens. The study also demonstrated the complexity of pest incursion pathways and population dynamics, especially in a connecting landscape with different national boundaries and separate biosecurity practices. Last but not least, a way to better understand biosecurity threats that have been impacting Brazil is to utilise museum samples of invasive species so as to build and track genetic diversity over time to better understand the country's long-term effort to protect the country's biological and agricultural resources from long-term biosecurity failures.

# SUMMARY

Rapid human population growth during the last two centuries has created significant challenge to agricultural production. Agricultural insect pests pose direct impact to the global agricultural production system, and when successful invasion into a new environment, will also likely pose significant biosecurity concern as potential invasive pests. Biosecurity measures therefore seek to protect the economy, environment and society from accidental or intentional introductions of invasive pests that can put the system at risk. More then ever, high agricultural productivity demands continued innovation to control invasive alien organisms such as weeds, diseases, insects, and other pests that are capable of evolving adaptive mechanisms such as resistance to different control measures, or as new species emerge or became dispersed to new regions. The vast majority of such species are necessary to maintain high agricultural productivity, but a small percentage of them can spread rapidly beyond their introduced areas and become invasive species. The biological invasions constitute a major driver to environmental change, affecting conservation, human health and agriculture.

Genetic tools if used correctly, can provide effective and rapidly confirmation of the presence/absence of certain invasive/exotic pest species so that potential economic losses may be avoided. Furthermore, they can assist with effective monitoring and management of exotic pest species once identified and appropriate markers developed. Managers seeking to control and/or to mitigate the spread of invasive pest species will benefit from information that helps to identify source populations and incursion routes. Effective genetic tools can also provide information on potential origin(s) of an invasive species, determining if introduction was intentional or as a result of unintentional released, and/or through escape from captivity. This may

also have implications for identifying the route of entry and assist in preventing further invasions from the same source.

Brazil has a high risk for the introduction and establishment of exotic insects because of its continental dimensions, large border region, and diverse climatic zones (equatorial, tropical and temperate). This work aims to fill some of these gaps in understanding biosecurity risks involving invasive pests in South America. The potential use of molecular tools to identify and study population genetics of two invasive pests in South America were explored. The aim was to provide early and quick species confirmation using NGS and molecular markers, and specifically to understand the genetic diversity of these invasive pest species, and to generate necessary knowledge to help improve the biosecurity in South America countries. Species used in this study were therefore chosen based on their economic impact in the regions where they were found.

The first study targets the Soybean Stem Fly (SSF) *Melanogromyza sojae*, which was found damaging soybean fields in southern Brazil. In the absence of specialists for identifying *Melanagromyza* spp. in Brazil, morphological identification was achieved by engaging Mr Hugh Brier, Senior Entomologist at the Queensland Department of Agriculture and Fisheries, Australia, prior to commencement of molecular characterization of this fly's mitochondrial DNA genome at the Commonwealth Scientific and Industry Research Organization (CSIRO) in Canberra, Australia.

The suspected flies specimens were identified by larval morphological characters as *M. sojae* (SSF) and the complete mtDNA genomes of three *M. sojae* collected from soybean (*Glycine max* L.) fields in Brazil were sequenced using the Illumina MiSeq NGS platform, and the specific mtDNA COI gene sequence was

characterised. The complete mtDNA genome also provided the opportunity to annotate all 13 protein coding genes of *M. sojae*, as well as the estimation of complete mitogenome sequence polymorphism rates. Robust DNA markers for *M. sojae* species identification and genetic studies were developed and a maximum likelihood phylogeny using partial mtDNA COI sequences of a representative individual was constructed to infer the phylogenetic position of the Brazilian *M. sojae*, and to determine if this fly species had previously been sequence characterised at the mtDNA COI gene region.

After the occurrence of SSF in Rio Grande do Sul and Santa Catarina states was confirmed, the second study aimed to investigate the population genetic diversity of SSF using the newly developed SSF-specific mtCOI molecular marker and discuss the potential biosecurity implications of incursions by *M. sojae* in Brazil. Brazilian *M. sojae* populations showed both high nucleotide diversity and high haplotype number in two preliminary and relatively small field surveyed samples, pointing to multiple founders with established populations. The study also identified a shared mtDNA COI haplotype with two *M. sojae* fly individuals previously sampled from Australia, and demonstrated the importance of integrating both traditional taxonomic research with NGS approach to unambiguously confirm the identity of an emerging global agricultural insect pest.

The global invasive lepidopteran pest *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae) was the target of the third and final study. This moth species' presence has been confirmed in North, South and Central America (in Brazil in 2013; in Paraguay in 2013; in Argentina in 2014 and in the USA in 2015), which has potential serious implications in terms of the management of insect pests in the main agricultural crops cultivated in these areas.

The genetic sequence data from the mtDNA COI partial gene of *H. armigera* specimens collected from Uruguay, Argentina, Paraguay, and also from three southern Brazilian states (Parana, Rio Grande do Sul and Santa Catarina) were analysed with the aim to better characterise the genetic diversity of *H. armigera* in the Cone Sul regions. The main results from this study were:

(i) the confirmation, at the molecular level, that *H. armigera* is present in soybean fields in Uruguay and Paraguay, and also in the Brazilian states of Rio Grande do Sul (RS), Santa Catarina (SC), and Parana (PR);

(ii) first molecular genetic diversity characterisation of *H. armigera* in the Cone Sul region (including individuals from Argentina);

(iii) detection of unexpected unique haplotype diversity in *H. armigera* populations from Uruguay, Argentina and Paraguay; and

(iv) findings that populations of H. armigera in the South American continent likely represented multiple independent incursion events. Findings that populations of *H. armigera* in the South American continent likely represented multiple independent incursion events will have significant implications to pest and resistance management strategies in the New World.

# SAMENVATTING

De snelle mondiale bevolkingsgroei van de afgelopen twee eeuwen heeft ons voor een grote uitdaging geplaatst op vlak van voedselproductie. Insectenplagen hebben een directe impact op het globale agriculturele systeem en zijn een grote zorg op vlak van bioveiligheid wanneer dergelijke plagen zich in nieuwe gebieden vestigen als mogelijke invasieve soort. Maatregelen op vlak van bioveiligheid zijn er daarom op gericht de economie, het milieu en de maatschappij te beschermen tegen ongewilde of opzettelijke introductie van invasive plagen die een risico kunnen vormen voor het ecosysteem. Meer dan ooit vergt een hoge agriculturele productiviteit continue innovatie om invasieve soorten zoals onkruiden, ziekten, insecten en andere plagen die in staat zijn om aanpassingsmechanismen te ontwikkelen, bijvoorbeeld resistentie tegen bestaande bestrijdingsmechanismen, het hoofd te bieden. De grote meerderheid van dergelijke soorten worden initieel ingezet om een hoge productiviteit te handhaven, maar een klein percentage kan zich snel verspreiden buiten de gebieden waarin ze geïntroduceerd worden, en zo een invasieve plaag worden. Deze biologische invasies spelen een grote rol in de verandering van ecosystemen en hebben een impact op de conservatie van inheemse soorten, op de menselijke gezondheid en de landbouw.

Genetische hulpmiddelen kunnen, bij correct gebruik, een effectieve en snelle bevestiging bieden omtrent de aan- of afwezigheid van bepaalde invasieve of exotische plaagsoorten, zodat potentiële economische verliezen vermeden kunnen worden. Bovendien kunnen ze ook bijdragen tot een effectieve monitoring en bestrijding van deze exotische plaagorganismen eens ze geïdentificeerd zijn en de correcte merkers ontwikkeld worden. Gewasbeschermers die op zoek zijn naar methoden om de verspreiding van invasieve plaagorganismen te beheersen of tegen te gaan, zullen

voordeel halen uit de informatie die bekomen kan worden omtrent de identificatie van de bronpopulatie en de verspreidingsroutes.

Effectieve genetische hulpmiddelen kunnen ook informatie verschaffen over de mogelijke herkomst van de invasieve soorten, om zo uit te maken of de introductie bewust was, of eerder het resultaat van een accidentele verspreiding of als resultaat van een ontsnapping. Ze kunnen ook helpen uit te maken hoe deze species zijn binnengebracht en kunnen zo helpen om verdere invasies in de toekomst te voorkomen.

Brazilië loopt door haar grote continentale oppervlak, lange grensregio's en diverse klimatologische zones (equatoriaal, tropisch en gematigd) een hoog risico wat betreft de introductie en vestiging van exotische insectsoorten. Deze thesis heeft als doel bij te dragen tot het beter begrijpen van bioveiligheidsrisico's op vlak van invasieve plagen in Zuid-Amerika. Het potentieel van moleculaire tools bij het identificeren van twee invasieve soorten in Zuid-Amerika en het bestuderen van hun populatiegenetica werd onderzocht. Het doel was om in een vroeg stadium en op een snelle manier bevestiging te kunnen geven omtrent de identiteit van de soort, met behulp van Next Generation Sequencing (NGS) en moleculaire merkers. Meer specifiek was het doel ook om de genetische diversiteit van deze invasieve soorten beter te begrijpen, en om de nodige kennis te genereren om de bioveiligheid in Zuid-Amerikaanse landen te verbeteren. De soorten die in deze studie werden gebruikt waren daarom gekozen omwille van hun economische impact in de regio's waarin ze werden aangetroffen.

De eerste studie handelde omtrent de kedelee-stengelboorder (KSB) *Melanogromyza sojae*, die werd aangetroffen in soja-velden in Zuid-Brazilië, waar het

insect voor aanzienlijke schade zorgt. Wegens de afwezigheid van specalisten in Brazilië, werd de morfologische identificatie van *Melanagromyza* spp. uitgevoerd door Dhr. Hugh Brier, entomoloog aan het Queensland Department of Agriculture and Fisheries in Australië. Vervolgens werd gestart met de moleculaire karakterisatie van het mitochondriale DNA genoom bij de Commonwealth Scientific and Industry Research Organization (CSIRO) in Canberra, Australië.

De identiteit van de verzamelde stalen werd bevestigd als zijnde *M. sojae* (KSB) en het complete mtDNA genoom van drie *M. sojae* individuen, verzameld in soja (*Glycine max* L.) velden in Brazilië werd gesequeneerd met behulp van het Illumina MiSeq NGS platform. Vervolgens werd het mtDNA-specifieke COI gen gekarakteriseerd. De kennis van de sequentie van volledige mtDNA genoom bood ook de mogelijkheid om alle 13 eiwit-coderende genen in het mtDNA te annoteren en om een schatting te maken van de hoeveelheid polymorfismen in het mitogenoom. Robuste DNA merkers voor *M. sojae* identificatie en genetische studies werden ontwikkeld en een maximum likelihood fylogenie, gebruik makende van partiële mtDNA COI sequenties van een representatief individu werd opgesteld om de fylogenetische positie van de Braziliaanse *M. sojae* af te leiden en om uit te maken of al eerder een sequentieanalyze voor het mtDNA COI gen van deze vlieg werd uitgevoerd.

Nadat de aanwezigheid van KSB in de deelstaten Rio Grande do Sul en Santa Catarina werd bevestigd, had de tweede studie tot doel de genetische diversiteit van de populaties KSB te onderzoeken, gebruik makende van de nieuw ontwikkelde KSB-specifieke mtCOI moleculaire merkers. Vervolgens werden de mogelijke implicaties op vlak van bioveiligheid en mogelijke introductie van *M. sojae* in Brazilië besproken. De

Braziliaanse *M. sojae* populaties vertoonden zowel een hoge nucleotide divergentie en een hoog haplotype aantal in twee preliminaire en relatief kleine veldstalen, wijzend op meerdere oorsprongshaarden bij de gevestigde populaties. Het onderzoek bracht ook een mtDNA COI haplotype aan het licht dat gedeeld wordt met twee *M. sojae* individuen die eerder in Australië werden verzameld en toonden het belang aan van het integreren van zowel traditioneel taxonomisch onderzoek en de NGS aanpak, om ondubbelzinnig de identiteit van een opkomende mondiale plaag aan te tonen.

De globale invasieve rupsenplaag *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae) was het onderwerp van het derde en laatste onderzoek. De aanwezigheid van deze mot werd reeds bevestigd in Noord-, Zuid- en Centraal-Amerika (in Brazilië in 2013, in Paraguay in 2013; in Argentinië in 2014 en in de Verenigde Staten in 2015). De aanwezigheid van deze plaag zorgt voor een verandering in prioriteiten en brengt zo potentieel grote gevolgen met zich mee wat betreft de controle van andere insect plagen die een bedreiging vormen voor de belangrijke gewassen die geteeld worden in deze streken.

De genetische sequentiedata van het partiële mtDNA COI gen van *H. armigera* specimen die verzameld werden in Uruguay, Argentinië, Paraguay en drie Braziliaanse deelstaten (Parana; Rio Grande do Sul; Santa Catarina) werden geanalyseerd met als doel de genetische diversiteit van *H. Armigera* in de Cone Sul (Zuidkegel) regio's beter te karakteriseren. De belangrijkste resultaten van dit onderzoek waren:

(i) De bevestiging, op moleculair niveau, dat *H. armigera* aanwezig is in soja-velden in Uruguay en Paraguay, alsook in de Braziliaanse deelstaten Rio Grande do Sul (RS), Santa Catarina (SC) en Parana (PR);

(ii) De eerste karakterisatie van de moleculaire genetische diversiteit van *H. Armigera* in de Zuidkegel van Zuid-Amerika (inclusief individuen uit Argentinië);

(iii) Detectie van onverwachte en unieke haplotypes in *H. Armigera* populaties uit Uruguay, Argentinië en Paraguay;

(iv) Bevindingen die aantonen dat de populaties van *H. Armigera* in Zuid-Amerika waarschijnlijk het gevolg zijn van meerdere onafhankelijke introducties van de soort. Deze bevindingen zullen significante implicaties hebben op vlak van plaagcontrole en resistentiebeheer strategieën in de Nieuwe Wereld.

**SUMÁRIO**

O rápido crescimento da população humana durante os últimos dois séculos criou desafios significativo para a produção agrícola. Insetos-praga de culturas agrícolas representam impacto direto no sistema de produção agrícola global, e quanto as pragas invasoras, uma invasão bem-sucedida em um novo ambiente representa uma preocupação de biossegurança significativa. As medidas de biossegurança buscam proteger a economia, ambiente e sociedade de introduções intencionais ou acidentais de pragas invasoras que podem colocar o sistema em risco. Mais do que nunca, as altas demandas de produtividade agrícola demandam inovação no controle de organismos exóticos invasores, tais como ervas daninhas, doenças, insetos e outras pragas que são capazes de desenvolver mecanismos adaptativos tais como a resistência a diferentes medidas de controle. Uma porcentagem dessas espécies pode se espalhar rapidamente para além das suas áreas introduzidas e se tornar espécie invasora. As invasões biológicas constituem um dos principais fatores para a alterações ambientais, afetando a conservação, a saúde humana e a agricultura.

Técnicas gnéticas se usadas corretamente, podem proporcionar confirmações efetivas e rápidas da presençca/ausencia de determinadas espécies praga invasoras/exóticas contribuindo para que potencias perdas economicas sejam evitadas.

Além disso, podem contribuir para o monitorizamento e gestão eficaz de espécies de pragas exóticas, uma vez identificados e marcadores apropriados desenvolvidos. Gestores que procuram controlar e/ou mitigar a propagação de espécies pragas invasoras irão se beneficiar de informações que ajudam a identificar populações de origem e rotas incursão. Ferramentas genéticas eficazes também pode fornecer informações sobre a potencial origem de uma espécie invasora, determinar

se a introdução foi intencional ou não intencional, e/ou por meio de escape de cativeiro. Isto também pode ter implicações para identificar a rota de entrada e auxiliar na prevenção de invasões adicionais da mesma origem.

O Brasil tem um alto risco para a introdução e estabelecimento de insetos exóticos devido as suas dimensões continentais, grande região de fronteira e diversas zonas climáticas. Esse trabalho visa preencher algumas das lacunas que existem na compreensão dos riscos de biossegurança envolvendo pragas invasoras na América do Sul. O potencial de uso de ferramentas moleculares para identificar e estudar a genética de populações de duas pragas invasoras na América do Sul foi explorado. O objetivo foi fornecer rápida confirmação de espécie utilizando NGS e marcadores moleculares e, especificamente, entender a diversidade genética dessas espécies pragas invasoras, gerando conhecimento necessário para melhorar as estratégias de biossegurança utilizadas nos países da América do Sul. As espécies utilizadas neste estudo foram, portanto, escolhidas com base em seu impacto econômico nas regiões onde foram encontradas.

O primeiro estudo teve como alvo a mosca da haste *Melanagromyza sojae*, a qual foi encontrada danificando lavouras de soja no sul do Brasil. Na falta de especialistas para identificar espécies de *Melanagromyza* spp. no Brasil, a identificação morfológica foi realizada com o suporte do pesquisador Hugh Brier, Entomologista Senior no Queensland Department of Agriculture and Fisheries, Australia, antes do início da caracterização molecular do genoma mitochondrial no CSIRO (Commonwealth Scientific and Industry Research Organization) in Canberra, Australia.

Os insetos suspeitos foram identificados por caracteres morfológicos larvais como *M. sojae* e os genomas mitocondrias completes (mtDNA) de três *M. sojae*

coletadas de soja (*Glycine max* L.) no Brasil foram sequenciados utilizando a plataforma Illumina MiSeq NGS, e a sequência específica do gene mitocondrial COI foi caracterizado. O anotação completa do mtDNA também proporcionou a oportunidade para anotar todos os 13 genes que codificam proteínas de *M. sojae*, bem como a estimativa das taxas de polimorfismo no mitogenome completo. Marcadores de DNA robustos para identificação de *M. sojae* e estudos genéticos foram desenvolvidos e uma filogenia de máxima verossimilhança utilizando sequências parciais mtDNA COI de um indivíduo representativo foi construído para inferir a posição filogenética do espécime brasileiro de *M. sojae*, e também para determinar se essa espécie de mosca já tinha tido essa região do gene mitocondrial (COI) previamente caracterizado.

Após confirmada a ocorrência de *M. sojae* no Rio Grande do Sul e Santa Catarina, o segundo estudo teve como objetivo investigar a diversidade genética da população de *M. sojae* utilizando o marcador molecular específico mtCOI-SSF recém-desenvolvido e discutir as implicações potenciais de biossegurança de incursões por *M. sojae* no Brasil. As populações brasileiras de *M. sojae* apresentaram tanto alta diversidade de nucleotídeos como elevado número de haplótipos nas duas amostras de campo preliminares e relativamente pequenas pesquisadas, apontando para vários fundadores com populações estabelecidas. O estudo também identificou um haplótipo COI-mitocondrial compartilhado com dois indivíduos de *M. sojae* amostrados anteriormente da Austrália, e demonstrou a importância de integrar a pesquisa taxonômica tradicional com abordagem de NGS para confirmar inequivocamente a espécie de um inseto praga emergente em nível agrícola global.

A espécie praga invasiva global *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae) foi o alvo do terceiro e último estudo. A presença desta espécie de

mariposa foi confirmada na América do Norte, Sul e América Central (no Brasil em 2013; no Paraguai em 2013; na Argentina em 2014 e nos EUA em 2015), trazendo sérias implicações em termos da gestão e manejo de inseto pragas nas principais culturas agrícolas cultivadas nessas áreas.

Os dados da seqüência genética de parte do gene mitocondrial COI de espécimes de *H. armigera* coletados do Uruguai, Argentina, Paraguai, e também de três estados do Sul do Brasil (Paraná, Rio Grande do Sul e Santa Catarina) foram analisados com o objetivo de melhor caracterizar a diversidade genética de *H. armigera* na região do Cone Sul. Os principais resultados deste estudo foram:

(i) confirmação, em nível molecular, de que *H. armigera* está presente em lavouras de soja no Uruguai e Paraguai, e também nos estados brasileiros do Rio Grande do Sul (RS), Santa Catarina (SC) e Paraná (PR);

(ii) primeira caracterização molecular da diversidade genética de *H. armigera* na região do Cone Sul (incluindo indivíduos provenientes da Argentina);

(iii) detecção de inesperada diversidade de haplótipos únicos em populações de *H. armigera* do Uruguai, Argentina e Paraguai; e

(iv) as populações de *H. armigera* no continente sul-americano provavelmente representam vários eventos de incursão independentes. Os resultados que as populações de *H. armigera* no continente sul-americano provavelmente são resultado de vários eventos de incursão independentes terão implicações significativas no manejo das estratégias de resistência dessa praga no Novo Mundo.

***CURRICULUM VITAE***

## Personal Information

**Full name** Jonas André Arnemann

**Birth** 09/08/1987 - Três Passos/RS - Brazil

**CPF Number** 016.672.910-84

---

### Formal Education

**2013** Doctorate in Agronomy.
Federal University of Santa Maria, UFSM, Santa Maria, Brazil
Gent University, Ghent, Belgium
Title: Molecular tools to identify and study invasive pests in Brazil
Advisor: Jerson Vanderlei Carús Guedes
Co- advisor: Guy Smaghhe
Co-advisor: Wee Tek Tay
Scholarship from: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

**2011 - 2013** Master's in Agronomy.
Universidade Federal de Santa Maria, UFSM, Santa Maria, Brazil
Title: Flutuação e densidade populacional de ácaros fitófagos associados a cultivares de soja no Rio Grande do Sul
Year of degree: 2013
Advisor: Jerson Vanderlei Carús Guedes
Co-orientador: Samuel Roggia
Scholarship from : Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

**2006 - 2011** Bachelor's in Agronomia.
Universidade Federal de Santa Maria, UFSM, Santa Maria, Brazil
Title: Reação de cultivares de soja ao ataque de ácaros fitófagos.
Advisor: Jerson Vanderlei Carús Guedes

Bachelor's incomplete in Filosofia.
Universidade Federal de Santa Maria, UFSM, Santa Maria, Brazil
Year of interruption: 2007

## Complementary Education

**2006** Estágio em Manejo Integrado de Pragas.
Universidade Federal de Santa Maria, UFSM, Santa Maria, Brazil

**2011 - 2011** Short Term Course in Programa ''De Olho no Futuro''.
Bayer CropScience, BAYER CROPSCIENC, Brazil

**2010 - 2010** Estágio Extracurricular - Março a Dezembro de 2010.
Fillmore Greenhouses, Inc., FILLMORE GREEN, United States

**2010 - 2010** Estágio Curricular - Janeiro a Março de 2010.
Fillmore Greenhouses, Inc., FILLMORE GREEN, United States

**2009 - 2009** Estágio Extracurricular - Jan a Fev de 2009.
Sementes Eliane SA, SEMENTES ELIANE, Brazil

**2009 - 2009** I Workshop sobre nível de controle de pragas/soja.
Embrapa Soja (Londrina - PR), EMBRAPA SOJA, Brazil

**2009 - 2009** Fundamentos de medidas físicas de solo e plantas.
Departamento de Solos da Universidade Federal de Santa Maria, DPTO DE SOLOS, Brazil

**2008 - 2008** Short Term Course in Planejamento de experimentos e redação científica.
Colégio Politécnico da Universidade Federal de Santa Maria, CPUFSM, Brazil

**2007 - 2007** Short Term Course in Poda das Plantas Frutíferas.
Colégio Politécnico da Universidade Federal de Santa Maria, CPUFSM, Brazil

**2006 - 2006** II Seminário sobre Tecnologia de Aplicação.
Bayer CropScience, BAYER CROPSCIENC, Brazil

**2006 - 2006** Short Term Course in Apresentação oral de trabalhos acadêmicos.
Programa de Educação Tutorial / Pet-Agronomia, PET-AGRO, Brazil

## S, T & A Production

### Bibliographic Production

**Articles Published in Scientific Journals**

1. **ARNEMANN,** J. A., WALSH, T., GORDON, K., BRIER, H., GUEDES, J. V. C., TAY, W. T. 2015. Complete mitochondrial genome of the soybean stem fly *Melanagromyza sojae* (Diptera: Agromyzidae). Mitochondrial DNA (Accepted 09-Sept-2015, In press)
2. **ARNEMANN**, J. A. ; FIORIN, R, L. ; PERINI, C. R. ; STORCK, L. ; CURIOLETTI, L.E. ; NACHMAN, G. ; GUEDES, J.V.C. 2015. Density and growth rates of spider mites in relation to phenological stages of soybean cultivars in Brazil. Experimental & Applied Acarology, 2015.
3. GUEDES, J. V. C. ; PERINI, C. R. ; STACKE, R. F. ; CURIOLETTI, L. E. ; **ARNEMANN**, J. A. ; ALENDE, V. P. . Lagartas da soja: das lições do passado ao manejo do futuro. Revista Plantio Direto, v. 144, p. 06, 2015.
4. **ARNEMANN, J. A.**, GUEDES, J. V. C., STACKE, R. F., MELO, A. A., PERINI, C. R., CURIOLETTI, L. E. Helicoverpa armigera: até no inverno. Cultivar Grandes Culturas. , v.183, p.26 - 28, 2014.
5. GUEDES, J. V. C., **ARNEMANN, J. A.**, PERINI, C. R., MELO, A. A., ROHRIG, A., STACKE, R. F., MACHADO, M. R. R. Helicoverpa armigera: da invasão ao manejo na soja. Revista Plantio Direto, v.137/138, p.24 - 35, 2014.
6. MELO, A. A., GUEDES, J. V. C., STORCK, L., SWAROWSKY, A., CAGLIARI, D., BURTET, L. M., **ARNEMANN, J. A.** Precipitação artificial após aplicação do inseticida clorantraniliprole associado com adjuvante em plantas de soja. Ciência Rural (UFSM. Impresso). , v.44, p.2118 - 2123, 2014.
7. GUEDES, J. V. C., **ARNEMANN, J. A.**, PERINI, C. R., MELO, A. A., ROHRIG, A. Helicoverpa armigera: Manejar ou perder. Cultivar Grandes Culturas. , v.176, p.12 - 16, 2013.
8. GUEDES, J. V. C., **ARNEMANN, J. A.**, STURMER, G. R., MELO, A. A., BIGOLIN, M., PERINI, C. R., SARI, B. G. Percevejos da soja: novos cenários, novo manejo. Revista Plantio Direto. , v.127, p.24 - 30, 2012.
9. GUEDES, J. V. C., **ARNEMANN**, J. A., BIGOLIN, M., PERINI, C. R., CAGLIARI, D., STACKE, R. F. Revisão necessária. Cultivar Grandes Culturas. , v.152, p.14 - 16, 2012.
10. GUEDES, J. C., KUSS-ROGGIA, R. C. R., STURMER, G. R., **ARNEMANN, J. A.** Percevejos da soja: proporção de espécies e distribuição espaço-temporal. Revista Plantio Direto. , v.108, p.15 - 21, 2009.
11. ROGGIA, S., GUEDES, J. C., KUSS-ROGGIA, R. C. R., **ARNEMANN, J. A.**, Návia, Denise. Spider mites associated to soybean in Rio Grande do Sul, Brazil. Pesquisa Agropecuária Brasileira (Online). , v.43, p.3 - , 2008.

## ACKNOWLEDGEMENTS

Firstly, I want to express my love and gratitude to my family, my dad Nelson, my mom Helena, my brother Jardel. I thank them for their strong support, incentive and encouragement for my education. I thank them for their love, care, prayers and bearing my absence during all these years, and still gave me all the emotional and financial support that I needed. My love and gratitude for them can hardly be expressed in words.

A special thank you to my fiancée, Cristiane Trivisiol da Silva. Words cannot describe how lucky I am to have her in my life. She has given more to me than I ever could have asked for. I love you, and look forward to our lifelong journey.

Also, thanks to my friend Cristiano De Carli, whom I consider as a brother, for his help and support.

I also would like to thank my promoter Jerson Vanderlei Carus Guedes, who received me in his LabMIP-Team as his graduation student at the Federal University of Santa Maria (UFSM) in 2007. Over the graduation, Masters and PhD courses, I have learned many things from him, with confidence and technical support during all these years.

Similarly, I would like to thank my supervisor from CSIRO, Wee Tek Tay, whom I am deeply indebted by the opportunity he gave for a fantastic learning experience at CSIRO, with experimental ideas shared by his knowledge, experience and scientific competence. It was a great privilege to be supervised by him. I thank him for his trust in my work and for all his support and patience during my time in Canberra. I also thank Wee Tek Tay for his care, and for offering me a place to stay in Canberra with his lovely family, Tammy, Ethan, Micah and Jesse. It was a pleasure for me to live with you for two months, you have made me felt at home!

Additionally, I would like to express my deepest appreciation to Tom Walsh, Karl Gordon, William James, and Melissa Piper, whom together with Tek, Dr Paul de Barro and Dr Garry Fitt, greatly supported my time at CSIRO. I thank them for sharing their knowledge. This work would certainly not be possible without them. I also thank Karl and Gisela for their hospitality and introducing me to the many wonderful sights and places around Canberra and New South Wales.

Special thanks to Mr Hugh Brier (Senior Entomologist, Queensland Department of Agriculture and Fisheries, Australia) for his assistance with the morphological identification of SSF, an essential step which permitted all subsequent research performed on this pest species.

I am very grateful to Dr Stephen Roxburgh (CSIRO Land and Water Flagship, ACT, Australia) for his enthusiasm and support for the haplotypes distributions analysis for *Helicoverpa armigera* data. Without his expertise and professionalism this work would otherwise not have been possible.

Thanks to Prof. Elgion Loreto and his team from LabDros-UFSM for all the support with DNA extraction, and sharing of technical support and knowledge.

I also recognize the importance to be surrounded by special colleagues and friends that I worked with. I express my gratitude to all my mates from LabMIP-UFSM for their friendship and assistance in my experimental setups. I also thank them for all the relaxing moments that we spent out of the lab, sharing too much fun. I particularly highlight the involvement of the Lab Tecnichian Gustavo Ugalde, who together with Ericmar Avila dos Santos, Gustavo Engel, Iuri Halmenschlager, Luis Eduardo Curioletti, Regis Felipe Stacke, Leonardo Burtet, Maiquel Pizzuti, Bruno Tomazzi, Alberto Rohrig, Darlei Bamberg, Adriano Arrue Melo, Clerison Regis Perini, Daniele Aguiar, Greissi

Giraldi, Natalie Feltrin, Thaiza Bass and Regina Stacke, helped solve many technical issues on this project.

I am grateful to Maicon Machado and Danilo Teles, for the figures they had helped prepared for Chapters II and III.

I also would like to thank all my colleagues and friends from the LabMIP-UFSM and from Laboratory of Crop Protection at the Ghent University.

I wish to also thank my PhD jury members: Prof. Dr. Rodrigo Jacques, Prof. Dr. Elgion Loreto, Prof. Dr. Juliano Boldo, Prof. Guy Smagghe, Prof. Peter Bossier and Dr. Wee Tek Tay, for reading and judging my PhD work. With their suggestions, this thesis was much improved.

Thanks also to Olivier Christiaens for helping with the Dutch summary translation and on previous RNAi experiments.

Thanks to Kevin Maebe for help with proofreading sections of the thesis draft.

Thanks also to the Federal University of Santa Maria, the CAPES Foundation, and CSIRO for the grants provided and the opportunity to carry out this research.

This project has been, without a doubt, the single greatest test of my own commitment, patience, and perseverance, spanning three countries in three continents. But I have not achieved this alone. Along the way, I have received support from too many people to count. Though you may not see your names here, in black and white, but know that your various contributions have not gone unnoticed or unappreciated.